DOCUMENT RESUME

ED 292 058 CS 009 061

AUTHOR	Zakaluk, Beverly L., Ed.; Samuels, S. Jay, Ed.
TITLE	Readability: Its Past, Present, and Future.
INSTITUTION	International Reading Association, Newark, Del.
REPORT NO	ISBN-0-87207-795-0
PUB DATE	88
NOTE	156p.
AVAILABLE FROM	International Reading Association, 800 Barksdale Rd., PO Box 8139, Newark, DE 19714-8139 (Book No. 795, $6.25 member, $9.50 nonmember).
PUB TYPE	Reports - Research/Technical (143) -- Information Analyses (070) -- Collected Works - General (020)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Decoding (Reading); Elementary Secondary Education; Foreign Countries; *Material Development; *Readability; *Readability Formulas; Reading Comprehension; Reading Research; Resource Materials; Textbook Evaluation; Textbook Research; Writing (Composition)
IDENTIFIERS	*Research Summaries; *Research Trends; Text Factors

ABSTRACT

Designed to appeal to students who want a short but up-to-date overview, researchers who are interested in a critical appraisal, and consumers who would like to know what leaders in the field think, this collection of articles highlights the changes that have occurred in readability research from the past to the present and makes predictions about the future. The articles and their authors are as follows: (1) "The Beginning Years" (Jeanne S. Chall); (2) "The Formative Years" (George R. Klare); (3) "Assigning Grade Levels without Formulas: Some Case Studies" (Alice Davison); (4) "Determining Difficulty Levels of Text Written in Languages Other than English" (Annette T. Rabin); and (5) "Writeability: The Principles of Writing for Increased Comprehension" (Edward B. Fry); (6) "New Ways of Assessing Text Difficulty" (Marilyn R. Binkley); and (7) "Toward a New Approach to Predicting Text Comprehensibility," (Beverly L. Zakaluk and L. Jay Samuels). (NH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Its Past,
Present,
& Future

2

READABILITY

Its Past, Present, and Future

Beverley L. Zakaluk
University of Manitoba

S. Jay Samuels
University of Minnesota

Editors

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ira

Published by the
International Reading Association
Newark, Delaware 19714

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L. W. Kline
(International Reading
Association)

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 309 061

INTERNATIONAL READING ASSOCIATION

OFFICERS
1987-1988

President Phylliss J. Adams, University of Denver, Denver, Colorado

Vice President Patricia S. Koppman, PSK Associates, San Diego, California

Vice President Elect Dale D. Johnson, Instructional Research and Development Institute, Boston, Massachusetts

Executive Director Ronald W. Mitchell, International Reading Association, Newark, Delaware

DIRECTORS

Term Expiring Spring 1988

Margaret Pope Hartley, The Psychological Corporation, North Little Rock, Arkansas
P. David Pearson, University of Illinois, Champaign, Illinois
Carol M. Santa, School District #5, Kalispell, Montana

Term Expiring Spring 1989

Marie C. DiBiasio, Bristol Public Schools, Bristol, Rhode Island
Hans U. Grundin, The Open University, Milton Keynes, England
Nancy E. Seminoff, Winona State University, Winona, Minnesota

Term Expiring Spring 1990

Jerome C. Harste, Indiana University, Bloomington, Indiana
Jane M. Hornburger, Brooklyn College, CUNY, Brooklyn, New York
Merrillyn Brooks Kloefkorn, Jefferson County Public Schools, Golden, Colorado

Copyright 1988 by the
International Reading Association, Inc.

Library of Congress Cataloging in Publication Data

Readability: Its past, present, and future / Beverley L. Zakaluk,
S. Jay Samuels, editors.
Includes bibliographies.
1. Textbooks – Readability. 2. Readability (literary style) 3. Reading comprehension. I. Zakaluk, Beverley L. II. Samuels, S. Jay. III. International Reading Association.
LB3045.8.R43 1988 371.3'2 – dc19 87-30549
ISBN 0-87207-795-0

Cover Design by Boni Nash

To

Lawrence B. Charry
1921-1982

Those who knew Larry Charry were saddened by his death in June 1982. He touched the lives of many people – a third grader who did one of his crossword puzzles, a student in one of his classes, a group of teachers whom he addressed.

A person of unlimited energy, Larry was an organizer, a planner, a leader. He was a teacher for thirty-three years. He taught reading/study skills in high school long before others recognized that need; he taught at West Chester University and Trenton State College. Larry investigated new ideas. He was interested in the use of computers in education, developed crossword puzzles to help students improve their reading comprehension, and he wrote and edited reading material for children. He felt that readability was an area neglected by many classroom teachers, and he founded the IRA Readability Special Interest Group to encourage the sharing of ideas and the dissemination of information.

Much more could be written about Larry and his accomplishments, but we could never describe adequately his ability to motivate people, his friendliness, his sincerity, or his enthusiasm for everything he approached. Larry is missed, but remembered.

John E. Boyd
St. Peters, Pennsylvania

Contents

Foreword *vii*

Preface *x*

Readability: Its Past

1 The Beginning Years 2
Jeanne S. Chall

2 The Formative Years 14
George R. Klare

Readability: Its Present

3 Assigning Grade Levels without Formulas: Some Case Studies 36
Alice Davison

4 Determing Difficulty Levels of Text Written in Languages Other than English 46
Annette T. Rabin

5 Writeability: The Principles of Writing for Increased Comprehension 77
Edward B. Fry

Readability: Its Future

6 New~~s~~ Ways of Assessing Text Difficulty 98
Marilyn R. Binkley

7 Toward a New Approach to Predicting Text Comprehensibility 121
Beverley L. Zakaluk and *S. Jay Samuels*

Contributors

Marilyn R. Binkley
Office of Educational Research and Improvement
Washington, DC

Jeanne S. Chall
Graduate School of Education
Harvard University
Cambridge, Massachusetts

Alice Davison
University of Wisconsin
Madison, Wisconsin

Edward B. Fry
Professor Emeritus
Rutgers University
New Brunswick, New Jersey

George R. Klare
Ohio University
Athens, Ohio

Annette T. Rabin
Lincoln University
Lincoln University, Pennsylvania

S. Jay Samuels
University of Minnesota
Minneapolis, Minnesota

Beverly L. Zakaluk
University of Manitoba
Winnipeg, Manitoba
Canada

Foreword

This book is organized around the theme of the past, present, and future of readability research. I shall draw on the articles in this volume to highlight the changes that occurred from the past to the present and will make some predictions about future readability research.

Past research in readability was atheoretical. Chall points out that the pioneers in the 1920s and 1930s tried numerous variables before discovering that sentence length and word difficulty were the best predictors of readability. They are not causes but indices of the semantic and syntactic difficulty of texts. Word frequency counts and readability criteria influenced basal readers from the 1920s to the 1960s when they used a whole word method of teaching beginning reading. In a review of research, Rabin demonstrates that readability formulas with language-required modifications have had a great impact throughout the world. Their impact is likely to continue until more appropriate and useful criteria displace the current, easy to use formulas.

An often cited misuse of readability formulas is the application of word and sentence length as writing criteria to make texts fit particular grade levels. Fry responds to this criticism by formulating readability inspired writing criteria to make texts comprehensible. His article might start a writeability research thrust.

Past readability criteria focused on the use of text characteristics to predict the grade at which an average reader would comprehend the text. It was assumed that the reader had the necessary resources for comprehension. The cloze technique does not predict and assume; it provides an actual try out on the material.

Researchers have questioned and experimented with the cloze technique word deletion rule for assessing comprehension. Binkley explains that some of the experimentation is to determine the text factors that influence learning and memory. Using cohesion analysis, she examines a text to identify the writer's cohesive style and then deletes cohesive ties to reflect this style. Analysis of the responses provides diagnostic information about the student – information on the student's intersentential integration ability and where the student should be placed on a reading development continuum.

In a novel approach to searching for alternatives to traditional readability criteria, Davison interviewed librarians and publishers of children's materials. She concludes that a set of principles can be developed for grading reading difficulty without using formulas for writing and editing.

Klare analytically reviews readability formulas from the past to the present. He concludes by pointing to Zakaluk and Samuels' method as the newest approach to readability.

Using an explicit theoretical formulation, Zakaluk and Samuels argue that a text's readability is a function of an interaction between text characteristics and reader resources. They then measure text difficulty and reader resources and insert this information into a nomograph to determine the readability of the text for individual readers. Their method is time consuming as they have to measure both text characteristics and reader resources (word recognition skills and prior knowledge for the text's topic). However, it is the beginning of research on a text-reader model of readability.

This book demonstrates that readability research has progressed from an atheoretical to a theoretical basis. In the past, researchers focused on text characteristics to predict readability. Currently, they are assessing text characteristics in interaction with reader resources. Researchers also are using readability procedures to test specific hypotheses derived from cognitive theory, thus attempting to find causal factors of readability.

Although predicting the future is hazardous, I think the interactive model will be a focal point of future research on readability. Researchers will continue to investigate hypotheses on the relationship between specific text factors and students' cognitive processes. The effects the teacher has on modifying text, enhancing reader resources, and establishing comprehension goals will enter into research on readability in a classroom setting.

This book should be appealing to students who want a short but up to date overview, researchers who are interested in a critical appraisal, and consumers who would like to know what leaders in the field think about the past, present, and future of readability research.

Harry Singer
University of California at Riverside

Preface

In 1982, while preparing for the 1983 IRA Convention, Lawrence Charry, founder and program chairperson of the Readability Special Interest Group, decided it was time to take a long look at readability–"past, present, and future." He stated, "Readability has been around fifty years, more or less. Has it worked? What are the benefits? What are the weaknesses?" He questioned also whether readability formulas were being used as they had been intended. He projected a careful look at the state of the art by those who could offer the greatest insight.

Although Dr. Charry did not live to see his plans come to fruition, a seminar, "Readability: An Historical Approach," was presented at the IRA Annual Convention in Anaheim in May 1983. Speakers at the seminar included Jeanne S. Chall, George R. Klare, Edward B. Fry, and S. Jay Samuels.

Recognizing the historical significance of the event, Samuels and Zakaluk volunteered to edit a monograph on readability in which the Anaheim papers would be included. They developed the framework for what follows, combining updated versions of the original papers with other manuscripts on related topics. Included in the latter group is information on readability research on text written in languages other than English, writeability, and nonconventional approaches to estimating text difficulty.

Since work was begun on this monograph, IRA and NCTE have issued a joint statement regarding the possible harmful effects of the uncritical use of readability formulas. This work was not intended as either an apology for or a defense of the use of readability instruments. It is a description of the status quo by those who are best qualified to give it. However, readability instruments would not have been developed had there not been a need for them, and it would be a great loss if their detractors were responsible for the cessation of current worldwide research. If we disregard all we have learned until now, we will be in the position of those in Venezuela as described by Nelson Rodriguez-Trujillo in the June/July 1985 issue of the IRA's *Reading Today*.

> In Spanish, we are at the other end of the spectrum. In this language, we confront the situation of having no readability formula good enough to explain even some of the variability of language difficulty and are suffering that absence. In evaluation committees, people argue and counterargue trying to decide whether materials are of a certain difficulty or appropriate for certain children. At the end, the issue is resolved on the basis of personal opinion, having in mind some abstract child. It is also a common situation to see teachers in the classroom selecting texts that are too difficult for children and forcing them to suffer frustration. The lack of information on readability of materials prevents teachers from responding to the students' different levels of ability.

The goal of this volume has been to review the field of readability and to suggest possible new directions. We hope we have been successful and that we have been able to clarify some of the issues, questions, and concerns readers might have regarding the use of formulas to evaluate written materials.

Annette T. Rabin
Lincoln University

IRA DIRECTOR OF PUBLICATIONS Jennifer A. Stevenson

IRA PUBLICATIONS COMMITTEE 1987-1988 James E. Flood, San Diego State University, *Chair* • James F. Baumann, Purdue University • Janet R. Binkley, IRA • Phyllis E. Brazee, University of Maine • Susan W. Brennan, IRA • Kent L. Brown Jr., *Highlights for Children* • Marie DiBiasio, Bristol Public Schools, Bristol, Rhode Island • Dolores Durkin, University of Illinois • Philip Gough, University of Texas at Austin • Margaret K. Jensen, Madison Metropolitan School District, Madison, Wisconsin • John Micklos, Jr., IRA • Ronald W. Mitchell, IRA • Joy N. Monahan, Orange County Public Schools, Orlando, Florida • Allan R. Neilsen, Mount St. Vincent University • John J. Pikulski, University of Delaware • María Elena Rodríguez, IRA, Buenos Aires • Jennifer A. Stevenson, IRA • Anne C. Tarleton, Albuquerque Public Schools, Albuquerque, New Mexico

The International Reading Association attempts, through its publications, to provide a forum for a wide spectrum of opinions on reading. This policy permits divergent viewpoints without assuming the endorsement of the Association.

Readability: Its Past

1

Jeanne S. Chall

The Beginning Years

The study of readability, in the sense of language comprehensibility, has a long history. It has deep roots in the classical rhetoric of Plato and Aristotle and in the vocabulary analyses of the Bible by ancient Hebrew scholars (Lorge, 1944). Although these ancient sources continue to enrich our understanding of language and text effectiveness, this chapter covers a shorter and more recent history of readability that began in the 1920s. It focuses on the contributions of educational researchers and on the use of readability in education. There is a considerable body of research and application of readability to general communication that, because of space limitations, cannot be covered here.

The Beginning Years of Readability Measurement

The beginnings of readability research came from two main sources – studies of vocabulary control and studies of readability measurement. Vocabulary control studies were concerned with the vocabularies that would be most effective for learning to read from reading textbooks. Specifically, they studied "new words" in each book, the number of times they were repeated, and their difficulty.

Readability measurement came from an interest in the comprehension difficulty of content area textbooks. During the beginning years, readability researchers devised procedures and instruments that would reliably and validly distinguish easier from more difficult texts or grade texts in order of difficulty.

Thus, the vocabulary control studies and the readability studies had similar purposes. Both sought objective means of measuring the difficulty of printed materials for learning and for comprehension. The vocabulary control studies concentrated mainly on primary level textbooks, while the readability studies were more interested in the comprehensibility of content texts and other materials written for students of middle and upper elementary grades, high school, college, and adults. Both areas of research started in the 1920s, although vocabulary control studies were more prevalent during the earlier years.

Both sets of investigations were concerned that the existing primary basal readers and content texts were too difficult for most students for whom they were intended. An early vocabulary study of fourth grade reading textbooks found a variation of from 20 to 40 percent in "unknown" words (Dolch, 1928). The impetus for the first readability study (Lively & Pressey, 1923) came from teachers who reported an unusual number of technical terms in junior high school science books, so that the study of the subject necessitated acquiring a scientific vocabulary rather than the learning of scientific facts and generalizations.

Thus, the first study of readability, similar to the early vocabulary control studies, was concerned not only with objective procedures for estimating difficulty but with making texts more readable for the students who used them. Why did this occur in the 1920s?

One factor was the publication in 1921 of the first extensive frequency word count of the English language, Thorndike's *Teacher's Word Book*, which provided an objective measure of word difficulty.

Another hypothesis is that the junior and senior high school population was changing in the 1920s. The population began to include more students who previously would have completed their formal schooling in the elementary grades. These "new students" were the first generation in their families to attend secondary school. Textbooks written for the earlier secondary school population, with stronger academic backgrounds, may have been too difficult for many of the newer secondary students.

It is harder to hypothesize why easier reading textbooks were sought for the primary grades, since compulsory schooling for elementary age children had been in effect for many years. One possibility is that more children were entering the elementary grades without the knowledge of English needed for reading existing textbooks. Interestingly, this is not mentioned in the early research literature on vocabulary control. A more likely hypothesis is the change in the 1920s from a heavier phonic to a heavier sight word approach for teaching beginning reading. The greater emphasis on sight recognition may have resulted in a need for lower vocabulary loads in reading textbooks, particularly in the primary grades. That this hypothesis has some validity is seen in the historical changes in the number of words in basal readers. Vocabulary counts decreased substantially from the 1920s to the 1960s, the years when sight word approaches were predominant. From the late 1960s to the early 1980s, the vocabulary loads of primary level basal readers increased considerably, as the amount of systematic instruction in phonics increased (Chall, 1983).

Research versus Mission

It is significant that the beginnings of both vocabulary control and readability measurement had their roots in changing social conditions. Researchers felt that easier textbooks would make students' learning more effective. Since the prevailing educational philosophy was to provide an education for all, researchers sought ways to assess whether this criterion was met. Thus, research started with the desire to find objective means to determine whether textbooks were suitable for those using them. Also, from the start, the work had a strong mission behind it – to use objective measures to select and produce textbooks suitable for *all* children.

This strong mission led to some unexpected outcomes. The early consensus that the books were too hard for most children, supported by the early data from various vocabulary and readability measures, led to recommendations that the books be made easier. But soon, recommendations for easier textbooks could not be supported by the research evidence.

In my 1958 review of the research on vocabulary control, I could find only one experimental study (Gates, 1930) designed to determine optimal difficulty of vocabulary for first grade reading. Gates tried experimental materials of varying vocabulary loads with children of different abilities. From these studies he estimated the number of repetitions necessary for best results with first grade children of different intellectual abilities. He found that 35 repetitions were best for those of average ability, 20 for those above average, and 40 to 45 for those below average.

When the vocabularies of basal readers are compared with the criteria established by Gates, we find that his standards were met in the late 1930s. By then, most first grade basals already had vocabulary loads recommended by the Gates experiment, but the vocabulary loads of primary basal readers continued to decline until the middle of the 1960s.

When there was so little research evidence on optimal vocabulary control, why did basal reader vocabularies continue to decline? There is probably no one answer, but several may have some validity.

There was confidence in the conclusions of the early vocabulary control studies that the basal readers were indeed too hard. (See Dolch, 1928.) There was also confidence that comparing the vocabularies of reading series from different publishers would lead to better standards of optimal difficulty. That is, an average from the various publishers would be closer to the optimal than the extremes. If most publishers followed this, it was inevitable that books would be easier with each new publication date.

Subject matter textbooks written for the higher grades, which were evaluated by readability measures during the beginning years, also were found to be too difficult for most students. And similar to vocabulary control researchers, readability researchers recommended easier textbooks. From the 1940s to the mid 1970s, there was a general decline in the difficulty of elementary and high school textbooks as judged by a variety of measures: readability formulas, level of maturity, difficulty of questions, and ratio of pictures to text (Chall, Conard, & Harris, 1977). After a long period of growing ease, reading instruction textbooks started to become more difficult in the early 1960s. This is explained best by the introduction of stronger phonics programs in most basal reading series in the late 1960s and the 1970s.

From the beginning years until the present, there has been a tendency to confuse the scientific study of vocabulary control and the measurement of readability with their educational uses. The scientific side of vocabulary control and readability measurement produced tools, procedures, and understandings that helped make possible optimal matches between readers and texts. Under certain conditions, this might suggest raising the level of difficulty; under other conditions, lowering the level. The objective tools made it possible to find that a book was too easy or too hard. But early research literature seldom reported that textbooks were too easy. This started to occur in more recent years.

Text Factors Studies in the Early Years

Vocabulary control, as well as early readability studies (1922 to 1926) tended to concentrate on vocabulary aspects such as difficulty, diversity, and range. The Thorndike frequency word lists or other word lists based on frequency of use in textbooks, readers, or by students in given grades were used to measure vocabulary difficulty. Judgment, experience, and correlational analysis were the standards for accepting one criterion of vocabulary difficulty as more reliable and valid than another.

During the early years of readability measurement, most researchers concentrated on vocabulary; in a second period of readability studies, investigation concentrated on a greater variety of factors (1928 to 1939). As early as 1926, the Winnetka Formula, designed to predict comprehension difficulty and interest in children's books, used several vocabulary and sentence factors (Washburne & Vogel, 1926). The end of this second period brought the Gray-Leary study (1935) that related eighty-two factors of text difficulty (including vocabulary, syntax, interest, and organization) to passages graded on the basis of reading comprehension performance by adults.

Standards of Optimal Difficulty

Generally, early studies of the vocabulary of reading textbooks found that basal reader series intended for the same grade varied widely in vocabulary difficulty and diversity and that most of the books used vocabularies outside the experience of children in the intended grades (Dolch, 1928). In most of these studies the conclusions were based on comparing the vocabularies of given basals with specific word lists, with the words in other basals, or with basals published at earlier dates. There were about seventy such studies during the late 1920s and early 1930s, and almost all referred to some word list as a basis for estimating whether a word was known to children in the grade, considering words outside the list unknown and hence difficult. Practically none of the studies actually tested the materials judged easier or harder.

Reliance on these standards continued into the 1940s. In 1941, Spache conducted an extensive analysis based on published vocabulary control studies, suggesting standards for the selection of primary level basal readers based on average vocabularies and ranges in published texts. On the assumption that most readers were still too hard, he indicated that the easier books for a given grade were "superior"; the harder books, "inferior."

What were the effects of these vocabulary studies? There is little doubt that the studies and their uses influenced authors of textbooks and publishers of educational materials. The studies also probably influenced state adoption committees and schools in the selection of textbooks and other instructional materials.

As noted earlier, the trend toward preferring reading textbooks with lower vocabularies came from the wide use of sight approach readers from the 1920s to the late 1960s (Chall, 1967, 1983). With less direct teaching of decoding and word recognition, the stories had to have fewer different words. Most early investigations recommended that the vocabularies of primary reading textbooks be low. There were some disagreements, however. Stone (1942) criticized a new reading series that presented only 1,147 different words through the third grade readers. And Yoakam (1945, p. 309) hoped that the use of readability formulas would "correct the tendency to make them too easy." The majority view seemed to win. The numbers and kinds of words in the basal readers continued to be an important issue in book production and selection. When books were ranked on various vocabulary factors, the results published, and superiority related to ease, publishers tried to meet or exceed the "averages." That this happened during the beginning years can be seen from a study by Hockett (1938), who found that first grade readers dropped from 644 different words in 1926 to 462 in 1937. Second grade readers dropped from 1,147 different words in 1930 to 913 in 1937.

The great drop in the vocabularies of primary readers from 1920 to 1960 (Chall, 1967, 1983) had its roots in several sources: the mission to teach all children, the early research results that concluded books were too hard, the changes in teaching methods, and the changes in student population. It also seems to have been influenced by research methods that unfortunately based their recommendations on comparisons and averages rather than on experimental tests of students and texts. Except for the Gates (1930) experimental study of optimal vocabulary size for first grade reading programs, there were no experimental studies to determine the best vocabulary standards for students of different abilities. The comparison of vocabularies to word lists in readers of the same grade but from different publishers was not enough. Empirical data were needed on the effects on children's learning as a result of using different vocabularies. With the consensus that easier is better, it was almost inevitable that readers became easier without strong evidence of their effectiveness.

Effects of Readability Assessment on Text Difficulty

What can explain the decline in content textbook difficulty during the early years and later? Several reasons seem valid. As noted earlier, there was a growing concern that schools must educate all students, particularly the "new" senior high pupils whose parents had less academic background compared with previous high school populations. The mission for easier books thus had some basis in reality. The existence of valid, reliable, and easy to use tools for estimating text difficulty levels gave further impetus to the use of readability measures by publishers, state and local textbook adoption committees, and schools. The growing concern for individualized instruction to meet the needs of students of varying reading abilities and the desire to match students' abilities to the difficulties of textbooks made the technology of the measurement more useful to publishers and teachers.

The growing ease of content textbooks, similar to the growing ease of reading textbooks, seems to stem from similar factors. Textbooks of the 1920s were probably too hard for the newer students, and there was a strong desire to make education more effective for them. What kept this mission so active and why did the textbooks become progressively easier, beyond that indicated by the research evidence and perhaps beyond their effectiveness for the students using them (Chall, Conard, & Harris, 1977)?

Teachers, publishers, text adoption committees, researchers, and the instruments themselves have been blamed for the growing ease of textbooks. The most recent tendency to blame word lists and readability formulas for the poor quality of textbooks is unfortunate; it is similar to blaming poor reading ability on the use of standardized reading tests. The causes are probably more complex and interrelated, but when understood will prove more helpful.

Optimal Difficulty

Perhaps the weakest aspect of readability and vocabulary research and their uses is the paucity of experimental studies to establish standards that are optimal for learning, comprehension, interest, and efficient reading.

The vocabulary and readability standards of the early years probably were headed in the right direction. The books probably were too hard for most of the students for whom they were intended. But continuing in that direction for forty or fifty years without experimental verification may have become dysfunctional. The curriculum changes, students change, teachers change, methods change, and expectations of achievement change. Thus, standards need continuous reevaluation.

The More Things Change

This section attempts to compare some of the concepts of readability and vocabulary measurement proposed during the beginning years with those of the 1980s.

As noted earlier, beginning readability research tended to focus on vocabulary and syntax, although investigations soon began to study other factors (Gray & Leary, 1935). In the 1980s, the readability concept of many tended to concentrate on text structure – organization, coherence, and cohesion.

Many researchers of the 1980s have been critical of the limited factors used in readability formulas. Newer critics give the impression that readability researchers in the early years overlooked factors other than words and sentences because they did not know they existed.

Excerpts from current analyses of readability and from researchers in the early years are presented here to give some insight into the historical development of the concept of readability.

Are readability measures concerned only with surface measures? An excerpt from Huggins and Adams (1980, p. 91) claims readability measures are concerned only with surface structures.

> Although readability measures can be found that correlate fairly well with text difficulty...their main weakness is that the difficulty of a passage involves its comprehension, and surface structure descriptions capture only some of the syntactic variables necessary to comprehension. As an extreme example of the inadequacy of these [readability] formulas, most of them would yield the same readability index on a passage if the word order within each phrase, and the order of the phrases within each sentence, were scrambled.

Ojemann (1934) indicates that in addition to vocabulary difficulty, composition, and sentence structure, such qualitative factors as concreteness or abstractness of relationships (as distinguished from individual words used), obscurity, and incoherence in expression were used in estimating text difficulty. The following excerpt from Ojemann reminds us of the excerpt from Huggins and Adams.

> In similar studies that have been carried out for the most part with school children, qualitative factors have been overlooked in general. Their importance may be made clearer by considering an extreme example. If in a set of paragraphs the sentences were arranged in random order, the number of sentences, the vocabulary difficulty, etc., would remain constant, but there is considerable possibility that comprehension would be interfered with (p. 19).

As is often assumed by researchers in the 1980s, Ojemann did not treat the hard words in a mechanical way. He noted that difficult passages contained difficult words because they discussed abstract ideas, and the easy passages used common words because they dealt with concrete experiences.

These excerpts from 1934 and 1980 make remarkably similar points. They say that we cannot look at readability factors in a surface or mechanical manner. Further, each reports that mixing sentences and words could give the same readability rating, but it would not be its true measure of comprehension difficulty.

The similarity of these observations raises an important issue with regard to earlier and later research in readability. Current researchers tend to view limitations as stemming from lack of knowledge. Yet much knowledge of language and communication with regard to text difficulty existed fifty or sixty years ago, possibly earlier. If the instruments and ideas were abused, explanations other than ignorance need to be sought.

Using Readability Measures for Writing and Rewriting

One current criticism of readability formulas is that they have led to poor quality writing because some editors and publishers have turned the readability formulas into means of obtaining lower readability scores. The claim is that writers use readability measures mechanically, substituting easier for harder words and shorter for longer sentences in order to achieve lower (assumed to be better) readability scores. The recent position paper on readability of the IRA and NCTE makes a point about this issue (Cullinan & Fitzgerald, 1984).

24

It is interesting to note that the current concern about the negative effects of mechanically simplifying texts also is expressed by readability researchers and scholars in the beginning years. In the 1930s Horn cautioned against the mechanical use of word lists and readability formulas for selecting and rewriting books in the social studies. He said that word lists and readability formulas do not adequately consider the conceptual difficulty of texts that may contribute to poor understanding, although the words may be common. Horn gave examples showing that words of high frequency may even cause greater difficulty since pupils may give words the wrong meanings. He further demonstrated from the studies of his students that negligible effects on comprehension may result merely from the simplification of vocabulary.

> There is a real danger that the mechanical and uncritical use of data on vocabulary will not only affect adversely the production, selection, and use of books but will result in absurdities that will throw research in this field into disrepute (Horn, 1937, p. 162).

The dangers that may stem from the use of readability formulas for mechanical rewriting of texts have always been of great concern, from the beginning years to the present. And the cautions from readability researchers to editors, publishers, and to schools that they should not use the formulas rigidly and mechanically also have been steady throughout the years. However, it was found that benefits for comprehension could be achieved when readability principles are used creatively (Chall, 1958).

To Conclude

Readability and vocabulary studies have a long history of research and application. At times we wonder why we seem to overlook the hard gained knowledge of the past in our eagerness to discover the new and why we seem to overlook the continuity of the old in the new. The present and future years in readability research and application should bring the knowledge of the early years to prominence again, while newer and better instruments and standards are developed. At the same time, it is hoped that the research on how best to use the instruments is kept current and in tune with the changing achievement of students, with standards, and with the knowledge and art of teachers and educational publishers.

References

Chall, J.S. *Readability: An appraisal of research and application.* Bureau of Educational Research Monographs, No. 34. Columbus, OH: Ohio State University Press, 1958. Reprinted by Bowker Publishing Company, Epping, Essex, England, 1974.

Chall, J.S. *Learning to read: The great debate.* New York: McGraw-Hill, 1967, 1983.

Chall, J.S., Conard, S., and Harris, S. *An analysis of textbooks in relation to declining SAT scores.* Prepared for the Advisory Panel on the Scholastic Aptitude Test Score Decline sponsored by the College Board and Educational Testing Service. Princeton, NJ: College Entrance Examination Board, 1977.

Cullinan, B., and Fitzgerald, S. *Statement on readability.* Joint statement by the Presidents of the International Reading Association and the National Council of Teachers of English, 1984.

Dolch, W.W. Vocabulary burden. *Journal of Educational Research,* 1928, *17,* 170-188.

Gates, A.I. *Interest and ability in reading.* New York: Macmillan, 1930.

Gray, W.S., and Leary, B.W. *What makes a book readable.* Chicago: University of Chicago Press, 1935.

Hockett, J.A. *The vocabulary and content of elementary school readers.* California Department of Education Bulletin No. 3. Sacramento, CA: State Department of Education, 1938.

Horn, E. *Methods of instruction in the social studies.* Report of the Commission on Social Studies, American Historical Association. New York: Charles Scribner's Sons, 1937.

Huggins, A.W.F., and Adams, M.J. Syntactic aspects of reading comprehension. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, NJ: Erlbaum, 1980.

Lively, B.A., and Pressey, S.L. A method for measuring the 'vocabulary burden' of textbooks. *Educational Administration and Supervision,* 1923, *9,* 389-398.

Lorge, I. Word lists as background for communication. *Teachers College Record,* 1944, *45,* 543-552.

Ojemann, R. The reading ability of parents and factors associated with reading difficulty of parent education materials. In G.R. Stoddard (Ed.), *Researches in parent education II.* Iowa City, IA: University of Iowa, 1934.

Spache, G. Problems in primary book selection of preprimers, primers and supplementary primers, first and second readers. *Elementary English Review,* 1941, *17,* January, February, April, May.

Stone, C.R. A vocabulary study based on 107 primary grade books. *Elementary School Journal,* 1942, *42,* 455.

Thorndike, E.L. *A teacher's word book of 10,000 words.* New York: Teachers College, Columbia University, 1921.

Washburne, C.W., and Vogel, M. *Winnetka graded book list.* Chicago: American Library Association, 1926.

Yoakam, G.A. The reading difficulty of school textbooks. *Elementary English Review,* 1945, *22,* 309.

2

George R. Klare

The Formative Years

Writers sometimes call text readable for rather different reasons. They may consider it *legible* or *interesting* or *comprehensible*. Over the years, *comprehensible* has become the most common reason. In fact, Harris and Hodges (1981) apply the terms *readability* and *comprehensibility* almost interchangeably in *A Dictionary of Reading and Related Terms*. This increased usage, at least in the field of education, stems from the widespread application of readability formulas. Many teachers think of readability primarily as a formula score.

A review of the formative years (up to the present) in the development of readability measures can add further background to Zakaluk and Samuel's presentation (Chapter 7) on the role of readability in matching materials to readers. By providing a review, this chapter serves as an aid to understanding present and future readability research and application and as an introduction to the ensuing chapters. The points listed, most typified by a readability formula, illustrate the developing concept of readability.

1. The almost exclusive emphasis on style variables in readability formulas.
2. The reduction of style variables to semantic and syntactic factors.
3. The search for a satisfactory criterion for formula development.
4. The presentation of readability formula scores in terms of grade levels.

5. The efficient use of a word list for the semantic factor and sentence length for the syntactic factor.
6. The efficient use of syllable length for the semantic factor and sentence length for the syntactic factor.
7. The trend to increased emphasis on ease of use.
8. The development of formulas for languages other than English.
9. The introduction of cloze procedure as a convenient criterion for formula development.
10. The growing criticism of readability formulas in terms of their developmental criteria and their grade level scores.
11. The growing criticism of readability formulas in terms of "writing to formula."
12. The need for improvement of current readability measures.

The history of readability is exhaustive. Chall (1958), Klare (1963, 1974), and Harrison (1980) provide added details for the earlier points. More recent points regarding research can be found in Klare 1982, 1984.

Almost Exclusive Emphasis on Style Variables

Gray and Leary did not develop the first readability formula; that distinction belongs to Lively and Pressey (1923). But their work and their influential formula (Gray & Leary, 1935) illustrate the first point exceptionally well. They began their research by collecting ideas about possible contributors to readability from 100 experts and 100 library patrons and put together a list of 289 so-called factors, which they grouped into four categories.

- Content
- Style of expression and presentation
- Format
- General factors of organization

Gray and Leary then cut this list to the 44 factors they could count reliably and which occurred often enough in the passages of their Adult Reading Test (their criterion passages) for statistical analysis. Of these 44 factors, 20 were significantly related to the scores of adults of limited reading ability. And of these 20 factors, they first used 8 in a multiple regression equation before finally settling on the 5 style factors in the formula below.

$$x_1 = -.01029x_2 + .009012x_5 - .02094x_6 - .03313x_7 - .01485x_8 + 3.774$$

x_1 = average comprehension score
x_2 = number of different hard words not on Dale List of 769
x_5 = number of personal pronouns
x_6 = average number of words per sentence
x_7 = percentage of different words
x_8 = number of prepositional phrases

Gray and Leary's formula yielded a multiple R of .65 with their criterion. They had hoped to include variables other than the five style elements they ended with, but others could not meet their requirements of being counted reliably, occurring often enough in their passages, and contributing sufficiently to their regression equation. As it happened, their procedure of combining only style variables in a regression equation became the typical pattern for formula development.

Reduction of Style to Semantic and Syntactic Factors

Washburne and Morphett were among the earliest of the formula developers; their first formula (Vogel & Washburne, 1928) came out shortly after Lively and Pressey's pioneer effort. Their second formula (Washburne & Morphett, 1938), however, shows the second point more clearly since the three variables they used reduce to semantic and syntactic factors.

$$x_1 = .00255x_2 + .0458x_3 - .0307x_4 + 1.294$$

x_1 = grade placement
x_2 = number of different words
x_3 = number of different uncommon words (outside Thorndike's 1,500)
x_4 = number of simple sentences in 75 sample sentences

Lorge's (1939) formula, which appeared soon after Washburne and Morphett's, illustrates the same point.

$$x_1 = .07x_2 + .1301x_3 + .1073x_4 + 1.6126$$

x_1 = grade placement
x_2 = average sentence length in words
x_3 = number of prepositional phrases per 100 words
x_4 = number of different hard words per 100 words not on the Dale 769 word list

Entin and Klare (1978a) factor analyzed three extensive readability matrices and found that semantic and syntactic factors still accounted for most of the variance. That is the good news; the bad news is that other kinds of style variables (literally hundreds have been tried) contribute so little added variance.

The formulas of Washburne and Morphett and of Lorge also illustrate the next point.

Search for a Satisfactory Criterion

Washburne and Morphett achieved a multiple R of .86 with their criterion, a remarkable value for its time. This might have been due at least partly to the nature of the criterion, which used the reading test scores of children who reported reading and liking a large sample of books. Few researchers have developed a criterion involving something other than comprehension alone, or one as extensive and labor intensive.

Lorge appeared to have found a more convenient criterion in McCall and Crabbs' *Standard Test Lessons in Reading* (1925). The Lessons had characteristics he and subsequent developers found very useful: a large number of passages, a variety of topics, a wide range of difficulty, and detailed grade levels. Lorge found a multiple R of .77 for his formula against this criterion. Though this value was somewhat lower than Washburne and Morphett's, the criterion was readily available and more convenient for statistical analysis, and *Standard Test Lessons in Reading* became the standard in early research. Lorge later discovered an error in the calculations for his formula that led him to publish a revised formula (1948). The revisions were slight—scores from the two formulas correlated +.94—but the new formula had a reduced multiple R of .67 with the McCall-Crabbs criterion passages. This might have been a cause for further research had it been discovered earlier, but the McCall-

Crabbs Lessons were too convenient for others to abandon (as will be noted later).

The formulas of Lorge and of Washburne and Morphett provided the standard illustrated in the next point.

Presentation of Scores in Terms of Reading Grade Levels

Readability formulas grew in popularity because they promised teachers a way of matching reading materials to the abilities of readers. With the earliest formulas, the match could not be made conveniently. The Washburne-Morphett and Lorge formulas, however, provided scores directly in terms of grade placement. This arrangement contributed to the popularity of formulas; in fact, certain formula makers added such scores to formulas that first appeared without them.

Use of grade level scores turned out to be a mixed blessing, since this practice contributed to disagreement among formula estimates. Formulas may disagree for other reasons, such as the variables used, the developmental criteria used, and the range of ability of the subjects. In addition, certain formula makers based their grade level criterion on the 50 percent comprehension level (e.g., where subjects could answer 50 percent of the questions on passages) and others on the 75 percent level. Thus, disagreements were bound to increase. McLaughlin (1969) insisted on the 100 percent comprehension level (whatever that is) in developing his formula. As a consequence, several writers found that McLaughlin's formula consistently gave reading level estimates about two grades higher than the widely used formulas.

The issue of grade level scores (particularly formula disagreements) continues, as will be emphasized later. Dale and Chall's formula (1948) illustrates both this point and the one to follow.

Efficient Use of a Word List and Sentence Length

Dale and Chall's formula was the most widely used formula in educational circles for many years. Their formula held up extremely well—in fact, surpassed others—in validity, and set the

Table 1
Correction Table for Use with the Dale-Chall Readability Formula

Formula Score	Corrected Grade Level
4.9 and below	4 and below
5.0 - 5.9	5 - 6
6.0 - 6.9	7 - 8
7.0 - 7.9	9 - 10
8.0 - 8.9	11 - 12
9.0 - 9.9	13 - 15 (College)
10.0 and above	16 + (College graduate)

stage for other word list formulas. The McCall-Crabbs passages served as criteria for developing this formula, presented here.

$x_{c50} = .1579x_1 + .0496x_2 + 3.6365$
x_{c50} = reading grade score of pupils who can answer correctly one-half the questions on a McCall-Crabbs passage
x_1 = percentage of words outside the Dale list of 3,000
x_2 = average sentence length in words

The formula used the 50 percent comprehension level as a criterion. However, a correction table also was provided for adjusting the formula scores to correspond more closely to difficulty, particularly for harder materials. (See Table 1.) Dale and Chall report that with these corrections the comprehension level falls between 50 and 75 percent.

The table also serves another useful correction purpose. As Bormuth (1966) pointed out, language variables do not necessarily relate to comprehension difficulty in a linear fashion, yet formula makers use linear equations. This can introduce a certain degree of error in the formulas. The matter of curvilinearity can be seen clearly in the difficulty scale itself when grade levels are used. As reading material at the higher grade levels (high school and college) is analyzed, the readability scores should begin to level off and go no higher than, perhaps, college graduate level. Throughout this part of the range, content knowledge increases in importance; readability formulas, being based on style variables alone, cannot adequately account for this. Furthermore, grade levels have little if any meaning beyond college graduate level.

Formulas that do not take account of curvilinearity can, at the extreme, provide absurd scores. An author submitted a passage from the California probate code for analysis by two research workers who had developed computer programs for one such formula. The programs were consistent—both reported that 122 years of schooling would be necessary to understand the passage!

The next formula, Flesch's popular Reading Ease, again illustrates the above point and the one to follow.

Efficient Use of Syllable Length and Sentence Length

Flesch had developed one readability formula prior to the publication of his later and much better known Reading Ease formula (1948). He first included a variable, called personal references, that he hoped would combine with his style difficulty variables, thus bringing interest value into the scores. Several research workers quickly pointed out that this served more to dilute than to strengthen the value of the scores. Consequently, Flesch proposed separate Human Interest and Reading Ease formulas, both using the McCall-Crabbs Lessons and the 75 percent comprehension level. As in similar attempts to bring in variables other than style difficulty, the former never achieved wide usage. The latter (below), however, became the most widely used formula outside educational circles.

RE = 206.835 - .846 wl - 1.015 sl
RE = Reading Ease, on a scale from 100 (very easy to read) to 0 (very difficult to read)
wl = average word length, in syllables
sl = average sentence length, in words

Flesch soon found that he needed a grade level scale to satisfy users. He went one better and provided two. The only difference in the two lay in the final columns of each. Both can be presented together in Table 2 with the last two columns the only ones not common to both.

Table 2
Interpretation Table for Flesch Reading Ease Scores

Description of Style	Average Sentence Length	Average Number of Syllables per 100 Words	Reading Ease Score	Estimated School Grades Completed	Estimated Reading Grade
Very Easy	8 or less	123 or less	90 to 100	Fourth Grade	Fifth Grade
Easy	11	131	80 to 90	Fifth Grade	Sixth Grade
Fairly Easy	14	139	70 to 80	Sixth Grade	Seventh Grade
Standard	17	147	60 to 70	Seventh or Eighth Grade	Eighth and Ninth Grades
Fairly Difficult	21	155	50 to 60	Some High School	Tenth to Twelfth Grades
Difficult	25	167	30 to 50	High School or Some College	Thirteenth to Sixteenth Grades (College)
Very Difficult	29 or more	192 or more	0 to 30	College	College Graduate

Flesch's interpretation tables once again provided for the curvilinearity in the grade level scale. In addition, the Reading Ease formula, being simple to apply, serves as a good illustration of the next point.

Trend to Increased Emphasis on Ease of Use

Danielson and Bryan (1963) were the first of many authors to develop a computerized readability formula to aid users in large scale applications. To make their programing simple, they used the variable of characters instead of syllables in their word count, and characters instead of words for their sentence count.

$\text{DB\#2} = 131.059 - 10.364\text{cpsp} - .194\text{cpst}$

DB#2 = score on a scale from 0 (hard) to 100 (easy)
cpsp = characters per space
cpst = characters per sentence

Syllables are much harder than characters to count by computer, but syllable counters have been developed by at least six authors. Similarly, computer programs have been developed for a large number of formulas. Schuyler (1982) published one of the best wide range programs in its entirety so that potential users could copy it. The program will handle nine different formulas for users.

Another significant move toward ease of usage is the Readability Graph developed by Fry (1963). The Graph permits a direct estimate of reading grade level on entering with syllable length and number of sentences per 100 word sample, thus providing another way of avoiding the manual use of a formula. It seems safe to say that in its most recent version (1977), Fry's Graph is the most widely used of all readability methods. The development of a hand calculator for it, and the surprising development of a parallel computer program for it, attest to its popularity.

Formulas for Languages Other than English

Work in the development of readability formulas for the English language began much earlier than similar work for other languages. Spaulding's (1951) formulas for Spanish as a second language were the first to be published. Following is the more commonly used of the three formulas eventually published.

$\text{Difficulty} = .1609(\text{asl}) + 33.18(\text{d}) + 2.20$

asl = average sentence length
d = density, based on a Density Word List

Despite the later start, much research on the readability of other languages has been published since 1951. Research workers have written at least eight books and have developed formulas for at least the following languages other than English.

Afrikaans
Chinese
Hebrew
Hindi (a modified American formula)

Danish
Dutch (several formulas)
Finnish
French (several formulas)
German (several formulas)
Korean
Russian
Spanish (many formulas)
Swedish
Vietnamese

Details of development work outside the United States can be found in Rabin's chapter in this publication.

Cloze Procedure as a Criterion for Formula Development

Cloze procedure—the deletion of words in text at stated intervals (usually every fifth word), which readers are asked to fill in correctly—can provide a good index of comprehensibility. This characteristic makes it a good potential criterion for the development of readability formulas. Though the cloze procedure was developed by Taylor in 1953, it was not until 1965 that Coleman first used it as a criterion. He developed four formulas. The one below, using the two variables found in so much recent research, yielded a multiple correlation of .89 with cloze criterion scores (adding more variables raised the correlation very little). Even more striking, the formula's cross-validation value reached .88.

$c\% = 1.16w + 1.48s - 37.95$
$c\%$ = percentage of correct cloze completions
w = number of one syllable words per 100 words
s = number of sentences per 100 words

Cloze procedure has several characteristics that soon made it a very popular criterion for formula development. It is objective in scoring, easy to use and analyze, uses the text itself as the test, and yields higher correlations than the McCall-Crabbs Lessons in comparison of the same formulas (Miller, 1972).

The popularity of cloze procedure was further enhanced when Coleman (1965; Miller & Coleman, 1967) and Bormuth (1969) developed extensive sets of passages scored in terms of cloze percentages correct. These passages have been used by others in developing and cross-validating their own formulas.

Criticism of Formulas in Terms of Developmental Criteria and Grade Level Scores

As noted earlier, McCall-Crabbs *Standard Test Lessons in Reading* proved to be a popular criterion for formula development, and such frequently used formulas as the Dale-Chall and Reading Ease were based on them. Consequently, questions about the appropriateness of the Lessons could inevitably raise questions about many formulas. Stevens (1980) raised just such questions. She quoted McCall as saying the Lessons were meant only to be practice exercises and were not intended for rigorous testing or criterion purposes.

What can be done in the face of such a charge? Research workers can take (and have taken) several different approaches.

1. Developers can turn to another criterion, such as a different reading test or cloze procedure. Cloze procedure has been the criterion of choice since well before the Stevens article. This is not to say that cloze is necessarily a perfect choice. Carver (1977) refers to it as a "rubber yardstick," since cloze scores reflect both the difficulty of the material and readers' ability. Kintsch (1979) considers it to be actually misleading as a measure of comprehension, arguing that it really is measuring redundancy instead.

2. Developers can restandardize the Lessons. Harris and Jacobson (1974, 1979) did this before Stevens' criticism, since they felt the earlier norms were out of date. They reported an earlier correlation of .74 for four variables, but later correlations in the high .80s and low to middle .90s, lending some support to this approach.

3. Research workers can examine comparable formulas based on the Lessons and on other criteria, to see how well the scores agree. In one such study (Klare, unpublished), three formulas that have the same index variables – word length in syllables and sentence length in words – were compared. The three formulas and their criteria were

- the Reading Ease formula (Flesch, 1948), based on the original Lessons norms;
- the Kincaid version of the Reading Ease formula (Kincaid et al., 1975), which used the Gates-MacGinitie Reading Test as the basis for its grade levels; and

- the Fry Graph (Fry, 1977), where the grade levels came (with some adjustment) from publishers' grade level assignments.

The examination was made across the range of grades, with the following results, suggesting that the Lessons and the formulas based on them may be more robust than Stevens' article suggests.

- The Flesch-Kincaid and Fry grade level assignments differed by no more than one grade in their common range of six to sixteen grade levels.
- Neither the Flesch-Kincaid nor the Fry differed by more than two grades from the Flesch Reading Ease assignments beyond these levels.
- The three formulas agreed (within one grade level) in their assignments for most grades.

Though this comparison showed a surprising amount of agreement, there is still the question of whether the formulas could simply be agreeing in giving incorrect grade levels. This question is difficult to answer satisfactorily, but Harrison (1980) has made a start in comparing the assignments of nine formulas against pooled teacher judgments. He used twenty-four first year secondary texts and sixteen fourth year secondary texts in British schools and found that the teachers' judgments yielded average reading age scores (reading grade level plus five) of 11.30 and 13.14 respectively. The two most predictive formulas, the Dale-Chall (1948) and the Mugford (1970), differed by half a year or less, up or down, from these values. The Reading Ease formula (Flesch, 1948) differed by one year; the FORCAST (Caylor et al., 1973) and SMOG (McLaughlin, 1969) differed by about two years; and the Fog Index (Gunning, 1952) differed by almost three years. More work of this sort could help teachers by telling them which formulas are most predictive. Another helpful approach would be regression equations permitting a user to relate grade levels from one formula to those of another.

4. Developers can abandon the use of grade level scores altogether, as recommended by a resolution of the Delegates Assembly of the International Reading Association (*Reading Research Quarterly*, 1981). This procedure has been followed by the College Entrance Examination Board (1980, 1982), which uses DRP (Degrees

of Reading Power) units for both its new test and its readability formula scores. The formula, a modification of one developed by Bormuth (1969), is presented below.

$$\text{R} = .886593 - .083640\,(\text{let}/\text{w}) + .161911\,(\text{dll}/\text{w})^3 - .021401\,(\text{w}/\text{sen}) + .000577\,(\text{w}/\text{sen})^2 - .000005(\text{w}/\text{sen})^3$$

R = readability in cloze units; this score is transformed into DRP units using the formula DRP = (1-R) x 100
let = letters in passage x
w = words in passage x
dll = Dale long list words in passage x
sen = sentences in passage x

The use of DRP units obviates the need for grade levels in both reading test and readability estimates (except in rough groupings for the selection of appropriate test forms). Careful modification of typical cloze techniques plus the use of the Rasch model also make it possible to avoid certain of the limitations of the cloze procedure in the preparation of parallel test forms. This approach is not without problems. Grade level scores, whatever their flaws, are familiar to teachers; DRP scores by themselves are not. Consequently, reading material cannot be matched to a particular reader's ability unless the reader has been tested with the DRP reading test and the material has been analyzed with the DRP readability formula. As a solution, the College Board has undertaken to analyze children's material with the DRP formula as it is published and to circulate a report (College Entrance Examination Board, 1982) on all such analyses. This is a big undertaking and still necessarily excludes old favorites published earlier; since the formula is too complex to apply easily by hand, software has been developed for Apple II computers. With the DRP arrangement, certain limitations inherent in degrees of comprehension can be overcome. For example, teachers can assign reading material at a reader's tested *independent* or *instructional* level and avoid material at *frustration* level.

The DRP program appears to be a significant step forward in matching reading material to readers. But it cannot answer all of the criticisms leveled at readability formulas, as the next point shows.

Figure 1
Two Ways of Looking at the Validity of Readability Measures

	Prediction of Readable Writing	Production of Readable Writing
Readability Variables	Index	Causal
Validity Check	Correlational	Experimental

Criticism of Formulas in Terms of Writing to Formula

Formula developers have long warned against the notion of writing to formula, arguing that it is at best misleading and at worst harmful. Klare (1976) has attempted to put the problem in perspective by distinguishing between two ways of looking at the validity of readability measures. Figure 1 provides a capsule comparison of the two.

Formulas can play a useful screening role in the prediction of readability, where only index variables in language are needed. But formulas cannot be used in the production of readable writing, because index variables are not sufficient for this purpose. Such use would be analogous to holding a match under a thermometer to warm a room. For producing readable writing, more variables must be considered in both the text and the reader. Davison (this volume) discusses this issue in detail along with the implications for textbooks and teaching materials. Fry (this volume) raises this issue again and discusses ways of writing more readably without misusing readability measures.

The following point returns to the readability measures themselves.

Need for Improvement of Readability Measures

In a recent article, Chall (1980) pointed to some educational problems surrounding the use of readability measures and argued for the need to improve current formulas. Others also have argued for improvement, notably Harris and Jacobson (1979). They point to the need to include variables other than style difficulty if readability is to move beyond Herbert Spencer, the English stylist of the past century.

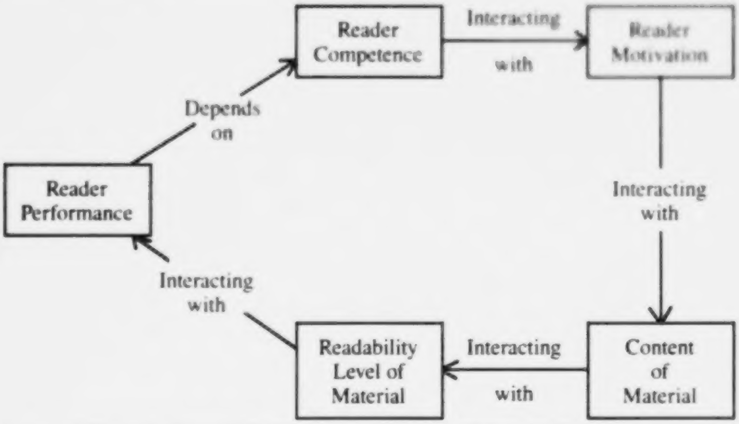
Klare (1976) has looked at the question of validity in a paper examining thirty-six experimental readability studies concerned with improving text comprehensibility. In nineteen of the studies, significant differences in comprehension were found; in eleven, the differences were not significant; in six, mixed results were found. The following categories of characteristics—28 variables in all—were examined in each study with a view to discovering which ones increased or decreased the probability of finding significant differences:

- experimental passages and how they were modified;
- tests and other dependent measures used;
- descriptions of the subjects and their characteristics;
- instructions given to the subjects;
- details of the experimental situation;
- statistical analyses employed; and
- results and detailed discussions based on them.

Figure 2, a slight revision (Klare, 1980) of the model in the earlier paper, summarizes the results.

Such a simple version of the model cannot adequately show the nature of the interactions or of the predictions that follow from the model, and space does not permit such detail here. It should be noted, however, that a number of experimental studies have supported predictions from the model. For example, Denbow (1973) compared easier and harder versions of two contents, one of higher and one of lower interest, with information gain as the dependent measure. He found that the easier of both contents produced significantly greater gain than the harder; however, the amount of gain was significantly greater on the content that was lower in interest value. Fass and Schumacher (1978) tested the effect of motivation directly by using two groups, one of which had a special monetary incentive. Once again, the easier version produced significantly greater comprehension only under lower motivation. Entin (Entin, 1980; Entin & Klare, 1985) provided further evidence of the effect of motivation by altering reader interest. She used twelve experimental passages,

Figure 2
A Model of Reader Performance


Reader Competence
Interacting with
Reader Motivation
Depends on
Reader Performance
Interacting with
Interacting with
Readability Level of Material
Interacting with
Content of Material

six shown previously to be of high interest and six of low interest to college freshmen. The passages were modified so that one version (standard) was at approximately grade twelve (freshman level) according to the Reading Ease formula and one was at approximately grade sixteen (difficult), yet with the same content according to judges. Both readability and interest resulted in a significant difference in cloze scores on the passages, i.e., an additive effect. She did not find the interaction effect found by Denbow and by Fass and Schumacher because the material was *at* (standard version) and *above* (difficult version) the readers' normal ability levels, so the effects of readability and interest were cumulative.

The degree of subjects' prior knowledge of content also can have an effect on whether readability changes will significantly affect comprehension scores. This was suggested in two earlier studies (Funkhouser & Macoby, 1971; Klare, Mabry, & Gustafson, 1955) but could only be presumed because of experimental conditions. Both studies seemed to show that as the degree of prior knowledge increased, the effect of readability decreased. In recent analyses, Entin and Klare (1978b, 1980) showed that a measured degree of prior knowledge had a clear effect. Correlations between readability levels and multiple-choice comprehension scores on test passages from a published reading test were essentially zero, but became moderately positive when corrections were made for prior knowledge of readers. Entin (Entin, 1980; Entin & Klare, 1985) later varied prior knowledge experimentally in her study of the effects of interest and readability and again found a significant effect for this variable. The effect was not as clear-cut as with interest and readability because of problems in getting a completely satisfactory measure of prior knowledge.

The above variables may interact with readability variables and thus play a part in whether formulas overestimate or underestimate reading difficulty. Can such reader variables be incorporated into readability formulas to improve their estimates? Not easily, but Kintsch (Kintsch, 1979; Kintsch & Vipond, 1979) has published some very encouraging results. Although arguing that he did not wish to "present a new readability formula," he reported a "proud .97" correlation for the following "formula" (Kintsch, 1979).

$$\text{Reading difficulty} = 2.83 + .48\text{rs} + .69\text{wf} + .51\text{pd} + .23\text{inf} + .21\text{c} - .10\text{arg}$$

Reading difficulty = number of seconds of reading time per proposition recalled on an immediate free-recall task
rs = number of reinstatement searches
wf = average word frequency
pd = proposition density
inf = number of inferences
c = number of processing cycles
arg = number of different arguments in a proposition list

In a later study (Miller & Kintsch, 1980), Kintsch found a multiple correlation of .86 between the same measure of reading difficulty on twenty passages by adding to the above the predictor variables of input size, sentence length, short term memory searches, and buffer size. His approach provides an interesting combination of traditional style variables with newer cognitive variables in achieving improved readability estimates, but it does require testing of potential readers. One can certainly hope that more labor intensive approaches such as this, the DRP method, or the estimation method described by Samuels and Zakaluk (this volume) will now be used. Perhaps a good way to put the matter is Rothkopf's comment (1980) that, for many, "practical considerations require the continued use of surface readability indicators" at this time.

In that case, the following summary suggestions might help to keep users of formulas from becoming misusers (Klare, 1984).

- Remember that different formulas may give different grade level scores on the same piece of writing. Though they may resemble thermometers in giving index values, they differ in that (like most educational and psychological tests) they do not have a common zero point.
- Look over existing formulas and pick a good one for the purpose at hand, but consider all formulas to be screening devices and all scores to be probability statements (Monteith, 1976).
- Choose a formula with two variables for rough screening purposes; having only one variable decreases predictiveness, but having more than two usually increases effort more than predictiveness. For critical applications or for research, apply more than one formula or try one of the newer, more complex formulas.
- Increase the value of an analysis by taking a large random (or numerically spread) sampling. For critical applications or for research, analyze the entire piece of writing (a computer program can be of help). For most books, three samples (often recommended) can give an indication of the average level of difficulty, but cannot say anything useful about variations in difficulty.
- Bear in mind that formula scores derive from counts of style difficulty; therefore, they become poorer predictors of difficulty at high grade levels (especially college) where content weighs more heavily.
- Consider again the purpose of the intended reading material; training readers calls for more challenging material than merely informing or entertaining them.
- Take into account other recognized contributors to comprehension; otherwise, formula scores may overestimate or

underestimate difficulty. For example, using special interests or incentives to get above average motivation can help to keep challenging material from being frustrating.

- Do not rely on formulas alone in selecting reading materials when this can be avoided. Include judges for characteristics that formulas cannot predict and to be sure that formulas have not been misused in preparing materials. Do not use just any judges – select experts or get more reliable opinions.
- Prepare to shift material after tentative placement. Where to draw the line between reading material that frustrates readers and reading material that challenges them cannot be specified easily with or without formulas.
- Keep formulas out of the writing process itself. If you use formulas for feedback, try the writing-rewriting cycle described by Macdonald-Ross (1979):

  Write ⟶ Apply formula ⟶ Revise ⟶ Apply formula....

No set of suggestions can prevent all formula users from becoming misusers. In the following chapter, Davison discusses this issue in greater detail.

References

Bormuth, J.R. Development of readability analyses. Final Report, Project No. 7-0052, Contract No. OEC-3-7-070052. Washington, DC: USOE, Bureau of Research, Department of Health, Education and Welfare, March 1969.

Bormuth, J.R. Readability: A new approach. *Reading Research Quarterly*, 1966, *1*, 79-132.

Carver, R.P. Toward a theory of reading comprehension and rauding. *Reading Research Quarterly*, 1977, *13*, 8-63.

Caylor, J.S., Sticht, T.G., Fox, L.C., and Ford, J.P. *Methodologies for determining reading requirements of military occupational specialties.* Technical Report No. 73-75. Presidio of Monterey, CA: Human Resources Research Organization, March 1973.

Chall, J.S. *Readability: An appraisal of research and application.* Columbus, OH: Bureau of Educational Research, Ohio State University, 1958.

Chall, J.S. Readability: In search of improvement. *Publishers Weekly*, October 29, 1980, 40-41.

Coleman, E.B. On understanding prose: Some determiners of its complexity. NSF Final Report GB-2604. Washington, DC: National Science Foundation, 1965.

College Entrance Examination Board. *Readability report: Degrees of reading power, 1982-1983.* New York: DRP Services, College Board, 1982.

College Entrance Examination Board. *User's manual: Degrees of reading power.* New York: DRP Services, College Board, 1980.

Dale, E., and Chall, J.S. A formula for predicting readability. *Educational Research Bulletin*, 1948, *27*, 11-20, 37-54.

Danielson, W.A., and Bryan, S.D. Computer automation of two readability formulas. *Journalism Quarterly*, 1963, *39*, 201-206.

Denbow, C.J. *An experimental study of the effect of a repetition factor on the relationship between readability and listenability.* Unpublished doctoral dissertation, Ohio University, 1973.

Entin, E.B. *Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages.* Unpublished doctoral dissertation, Ohio University, 1980.

Entin, E.B., and Klare, G.R. Components of answers to multiple-choice questions on a published reading comprehension test: An application of the Hanna-Oaster approach. *Reading Research Quarterly*, 1980, *15*, 228-236.

Entin, E.B., and Klare, G.R. Factor analyses of three correlation matrices of readability variables. *Journal of Reading Behavior*, 1978a, *10*, 279-290.

Entin, E.B., and Klare, G.R. Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages. In B. Hutson (Ed.), *Advances in reading/language research*, volume 3. Greenwich, CT: JAI Press, 1985.

Entin, E.B., and Klare, G.R. Some interrelationships of readability, cloze, and multiple-choice scores on a reading comprehension test. *Journal of Reading Behavior*, 1978b, *10*, 417-436.

Fass, W., and Schumacher, G.M. Effects of motivation, subject activity, and readability on the retention of prose materials. *Journal of Educational Psychology*, 1978, *70*, 803-808.

Flesch, R.F. A new readability yardstick. *Journal of Applied Psychology*, 1948, *32*, 221-233.

Flesch, R.F. The art of readable writing. New York: Harper & Row, 1949, 177.

Fry, E.B. Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading*, 1977, *21*, 242-252.

Fry, E.B. *Teaching faster reading.* London: Cambridge University Press, 1963.

Funkhouser, A.R., and Macoby, N. *Study on communicating science information to a lay audience*, phase 2. Report NSF 92-996. National Science Foundation, Institute for Communication Research, Stanford University, 1971.

Gray, W.S., and Leary, B.E. *What makes a book readable.* Chicago: University of Chicago Press, 1935.

Gunning, R. *The technique of clear writing.* New York: McGraw-Hill, 1952.

Harris, A.J., and Jacobson, M.D. A framework for readability research: Moving beyond Herbert Spencer. *Journal of Reading*, 1979, *22*, 390-398.

Harris, A.J., and Jacobson, M.D. *Revised Harris-Jacobson readability formulas.* Paper presented at the annual meeting of the College Reading Association, Bethesda, Maryland, October 1974.

Harris, T., and Hodges, R. *A dictionary of reading and related terms.* Newark, DE: International Reading Association, 1981.

Harrison, C. *Readability in the classroom.* Cambridge: Cambridge University Press, 1980.

Kincaid, J.P., Fishburne, R., Rogers, R.L., and Chissom, B.S. *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy enlisted personnel.* Branch Report 8-75. Millington, TN: Chief of Naval Training, 1975.

Kintsch, W. On modeling comprehension. *Educational Psychologist*, 1979, *14*, 3-14.

Kintsch, W., and Vipond, D. Reading comprehension and readability in educational practice and psychological theory. In L.G. Nilsson (Ed.), *Perspectives on memory research.* Hillsdale, NJ: Erlbaum, 1979.

Klare, G.R. *A manual for readable writing*, fourth edition. Glen Burnie, MD: REM, 1980.

Klare, G.R. A second look at the validity of readability formulas. *Journal of Reading Behavior*, 1976, *8*, 129-152.

Klare, G.R. Assessing readability. *Reading Research Quarterly*, 1974, *10*, 62-102.

Klare, G.R. Readability. In H.E. Mitzel (Ed.), *Encyclopedia of educational research*, fifth edition. New York: Free Press, 1982, 1520-1531.

Klare, G.R. Readability. In P.D. Pearson (Ed.), *Handbook of reading research*. New York: Longman, 1984, 681-744.

Klare, G.R. *The measurement of readability*. Ames, IA: Iowa State University Press, 1963.

Klare, G.R., Mabry, J.E., and Gustafson, L.M. The relationship of style difficulty to immediate retention and to acceptability of technical material. *Journal of Educational Psychology*, 1955, *46*, 287-295.

Lively, B.A., and Pressey, S.L. A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision*, 1923, *9*, 389-398.

Lorge, I. Predicting reading difficulty of selections for children. *Elementary English Review*, 1939, *16*, 229-233.

Lorge, I. The Lorge and Flesch readability formulae: A correction. *School and Society*, 1948, *67*, 141-142.

Macdonald-Ross, M. Language in texts. In L.S. Shulman (Ed.), *Review of research in education*. Itasca, IL: Peacock, 1979.

McCall, W.A., and Crabbs, L.M. *Standard test lessons in reading*. New York: Bureau of Publications, Teachers College, Columbia University, 1925.

McLaughlin, G.H. SMOG grading: A new readability formula. *Journal of Reading*, 1969, *12*, 639-646.

Miller, G.R., and Coleman, E.B. A set of thirty-six passages calibrated for complexity. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 851-854.

Miller, J.R., and Kintsch, W. Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, *6*, 335-354.

Miller, L.R. *A comparative analysis of the predictive validities of four readability formulas*. Unpublished doctoral dissertation, Ohio University, 1972.

Monteith, M.K. Readability formulas. *Journal of Reading*, 1976, *19*, 604-607.

Mugford, L. A new way of predicting readability. *Reading*, 1970, *4*, 31-35.

Resolutions passed by Delegates Assembly. *Reading Research Quarterly*, 1981, *16*, following page 613.

Rothkopf, E.Z. Copying span as a measure of the information burden in written language. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 562-572.

Schuyler, M.R. A readability program for use on microcomputers. *Journal of Reading*, 1982, *25*, 560-591.

Spaulding, S. Two formulas for estimating the reading difficulty of Spanish. *Educational Research Bulletin*, 1951, *30*, 117-124.

Stevens, K.C. Readability formulae and the McCall-Crabbs Standard Test Lessons in Reading. *The Reading Teacher*, 1980, *33*, 413-415.

Taylor, W.L. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, *30*, 415-433.

Vogel, M., and Washburne, C. An objective method of determining grade placement of children's reading material. *Elementary School Journal*, 1928, *28*, 373-381.

Washburne, C., and Morphett, M.V. Grade placement of children's books. *Elementary School Journal*, 1938, *38*, 355-364.

Readability: Its Present

3

Alice Davison

Assigning Grade Levels without Formulas: Some Case Studies

Readability formulas have been widely used to assign grade levels to texts on the basis of two text properties – average sentence length and average word complexity. Since the formulas' development in the 1920s and 1930s, reading researchers have been aware of their limitations for assigning accurate and meaningful grade levels (Gray & Leary, 1935). Yet formulas continue to be used, particularly for assigning difficulty levels in school textbooks, because there are no simple, convenient alternatives that would assign more accurate levels.

For the same reason – that there is no obvious alternative – formulas continue to be used for another, less justifiable, purpose. Texts often are edited to reduce their readability by simplifying vocabulary and shortening sentences. In the process, comprehensibility is not improved, while explicit connections as well as expressive and interesting words are lost. This fact about adaptations used in basal readers has been noted and documented many times. Ohanian (1987) describes how much of an interesting story is lost when the syntax and vocabulary are simplified to meet the readability level assigned to a basal reader.

Some problems with formulas are reviewed in other chapters of this book, and other questions can be raised about the validity of using formulas to predict whether a particular text can be read by a specific reader or group of readers. Another problem is that formu-

Research supporting this chapter was provided under Contract No. 400-81-0030, U.S. Office of Educational Research, to the Center for the Study of Reading, University of Illinois.

las are based on correlations with factors that cause comprehension difficulty, not on the actual causes of difficulty. These issues are reviewed in more detail in Davison and Green, 1988. Since formulas do not define the sources of difficulty, they cannot be used as guidelines for writing. The guidelines (including clear organization, explicit connectives, and appropriate vocabulary) that have been suggested are too general or subjective to be alternatives for formulas (Davison & Kantor, 1982).

Readability formulas probably will continue to be used until some widely applicable alternatives are found. The purpose of this chapter is to give case studies of alternative procedures that already exist for assigning grade levels. These case studies describe some situations in which it is not possible to use formulas and others in which reviewers or editors have chosen not to use formulas because of their many drawbacks. From these case histories and other similar situations, it may be possible to discover generally applicable alternatives to readability formulas.

Trade books for children

The term *trade books* refers to books intended for children to read outside of school in their leisure time. Teachers, librarians, and parents often need to know which of these books would be appropriate (in terms of reading difficulty) for a particular child or group of children. Compared with school texts, trade books vary a great deal in terms of subject matter, style, presentation, vocabulary, and sentence structure. Many of the factors that influence whether a reader will find such books difficult or easy to understand cannot be measured by formulas. Rather, a skilled and experienced teacher or other authority must estimate the age range and reading ability appropriate for a particular book.

Specialists in children's literature, or librarians who are familiar with children's reading preferences, can read a book analytically and judge accurately its probable difficulty level without using the word and sentence counts that go into formulas. The levels reviewers assign are more relative and flexible than the levels assigned by formulas. Reviewers' levels cover a two to three year range for average readers in those age groups. But children in lower grades may still like the book if they are very skilled readers, while somewhat older children may enjoy the book if they do not read as well as average children in their age or grade level.

Some of the factors reviewers consider are writing style (use of unusual words or complex sentence structures), the overall organization of the book, and the kind of exposition used. In general, children like very clear organization, with the episodes following a normal sequence of time or progression of ideas from simple to more complex. The characters in a story also influence children's responses, since children tend to identify with protagonists of their own age or slightly older.

These factors are not the only criteria for judging a book's reading level. Much depends on the individual reader. For example, poor readers find difficult words a great obstacle to reading, while average to good readers do not have difficulty in understanding texts because of such words. Rather they find unusual, expressive, or colorful words amusing and interesting. An unusual kind of exposition, such as one using flashbacks, can be made clear and interesting by a skilled author who uses it to heighten suspense or create an atmosphere of mystery. Young children may like certain stylistic features, such as plays on words, that older children might find silly.

Some confirmation of the accuracy of estimated grade and age levels is usually available. Well written books that are appropriate for a particular level in content and style become successes. They are borrowed frequently from libraries and are kept in print for a long time. Therefore, trade books provide a natural laboratory for discovering what makes books accessible to children.

Interestingly, trade books were the basis for the first readability formulas (Vogel & Washburne, 1928) to come into general use. Vogel and Washburne tested a large number of children for reading ability on a standardized test. They also asked the children for titles of books they had read and liked. Vogel and Washburne correlated features of these books with the reading level of the readers who mentioned them. Similar studies could be done now to define what properties of trade books make them accessible and popular. The result could be to define operationally how reviewers assign grade levels. This procedure could give some insight into the features of texts that make books comprehensible to different age groups.

Trade books are an important example of a situation in which readability formulas cannot be used accurately to judge text difficulty. Trained, experienced adult readers weighing a number of relative factors can successfully assign grade and age ranges to trade books. Furthermore, it is possible to validate these subjective judgments when a book becomes a favorite with children at a certain grade level.

Science magazines for children

Like trade books, magazines for children on such topics as nature, science, and exploration are a natural laboratory for discovering what features need to be considered in writing readable texts. Since editors and publishers of these magazines want children to understand and enjoy the articles, they must decide how to present information in an appropriate way for the intended audience. One of the problems they face is that articles on scientific concepts tend to use complex, technical words. An explanation of a scientific idea may have to relate ideas in complex and often long sentences. Each of these factors would increase the readability level assigned by readability formulas, perhaps in a way that does not reflect the actual difficulty level of the text. That is, a formula may not be sensitive to real obstacles to comprehension in one text, while at the same time it predicts a high level of difficulty for another text that is actually quite clear in most ways.

To see what alternatives to formulas might exist, we can compare similar (and equally popular) science magazines. The magazines are different in that Magazine A does not use readability formulas at all, while Magazine B uses formulas extensively for editing the stories in each issue.

Both magazines try to choose stories that have subject matter interesting to children. They also try to limit the length of stories and to leave out topics that cannot be expressed clearly within these limits. They use illustrations to arouse interest in reading the article and to focus the reader's attention on important ideas.

The basic difference in approach lies in the way articles are planned and edited. Magazine B, which uses readability formulas, starts with the picture layout of the page. Space is assigned for an article of a certain length, which is then written to fit this space. As much as possible, the article has to explain or refer to the pictures in the already existing layout. If the article has too many difficult words, or exceeds a certain level (sixth grade on the Fry formula), it must be rewritten. Rewriting often means words are simplified and sentences are shortened. The result in some cases is that the article cannot be structured to explain the subject matter clearly, and the relation between the pictures and the text is not always apparent.

Magazine A has a policy of not using readability formulas. The articles are planned to appeal to second and third grade children who are beginning to read on their own, fourth and fifth grade children who can understand somewhat more complex articles, and fifth and sixth grade children who can read even more complex stories. Writers are given a set of guidelines for presenting the subject matter at one of the age and grade ranges between second and sixth grade.

A closer look at the stories in Magazine A shows how the writers and editors match the text with the intended readers. The fact that readability formulas are not used can be confirmed by looking at the average sentence length and use of words in the articles. The average length of sentences in all the articles does not vary much, regardless of the intended age and grade level. The length of the articles themselves does vary, with the shortest ones being intended for the youngest readers. The use of conjunctions like *when*, *if*, and *because* increases in articles for older and more skilled readers. What distinguishes the levels of difficulty is the content and presentation of the subject matter.

The selections for the second and third grade children are short, rely heavily on pictures, and are usually about a young animal. Different episodes present the animal in a way that young children can easily identify with. In a clear time sequence, the article describes relationships with parents and siblings and how the animal eats, sleeps, is protected from danger, and learns new skills.

The articles for grades three to five tend to focus on classes of creatures that are different but have some common characteristic, such as living in water or sharing a particular habitat. These articles emphasize contrasts and similarities as well as the relationship between an animal and its environment. Articles of this type build on concepts about individual species familiar to younger children, but teach the readers to see more abstract generalizations about individuals that do not look alike.

The articles for children in grades five and six are longer and refer to more abstract or complex concepts, including cause and effect. Some of the articles introduce the idea of theories and hypotheses, intended to explain known and observed facts. Through reading about how scientists form hypotheses to explain natural phenomena and how these hypotheses are tested, children learn scientific reasoning and how to evaluate an explanation in relation to the evidence for it.

In all the articles in Magazine A, the topic is made clear in the first part of the article, and the presentation of ideas follows a clear pattern. In stories for younger children, the sequence is usually chronological, without flashbacks. In articles for older children, ideas are presented in a logical order, either following temporal sequence or placing cause before effect. What is not found in Magazine A, but is common in Magazine B, is the organization of information common to newspaper articles. This organization places the most important or striking facts first, the next most important ones second, and so on, which tends to make the connections between ideas less clear, especially to younger readers.

Publishers of children's magazines face the same problems as publishers of science or social studies textbooks. They must convey complex, abstract ideas to children with limited conceptual knowledge and reading ability. Readability formulas have limited applicability to texts on these topics. An alternative set of principles to use in writing scientific material for children can be derived by comparing and analyzing selections found in successful publications that do not use readability formulas. These guidelines include attention to interest, overall length, organization, and method of exposition.

They also include the careful use of appropriate illustrations and a choice of topics appropriate to specific age levels. Consequently, publishers are able to provide articles that are graded for conceptual difficulty and do not exceed the reading capacity of the readers. These goals can be accomplished without using formulas for writing and editing.

Languages Other than English

To estimate the difficulty of a text in another language also requires devising alternative techniques. In formulas developed for English, correlations between comprehension performance and word difficulty and sentence length are based on texts written in English and do not automatically carry over to other languages. Many languages have word and sentence structures different from those found in English. Although a formula could be adapted for another language, with changes reflecting what is difficult or complex, the revised formula could not be used reliably without being validated for texts in the new language. Such a procedure requires a substantial investment of time, effort, and money. Instead of adapting formulas, educators and researchers have tried to go directly to the issues involved in text difficulty. What they have done shows another approach to matching students with texts without using formulas.

In India, researchers are trying to develop reading achievement tests for seventh grade students in the major regional languages, which are languages in which reading instruction is given. It is not possible, however, to construct a test of this kind without knowing which texts require a particular level of reading ability. One such language is Telugu, a South Indian language used in Andhra Pradesh, one of the states of India.

It would not be practical to try to adapt an English readability formula for use on Telugu texts. The correlation in English between familiar, easy words and words of one or two syllables does not hold in Telugu. Nouns and verbs may have multisyllabic affixes for case or tense endings. The more difficult words generally are not longer than familiar words. The sentence structure of this language is more like Japanese than English, and for many reasons the correlation in English between long sentences and complex sentences does not carry over to Telugu. For example, a sentence with subordinate clauses can be the same length as a simple sentence.

So, to estimate levels of text difficulty, Indian government researchers are relying on the judgment of teachers who have taught at the seventh grade level and are familiar with what kind of texts can be read by students who are making good progress in reading. A certain number of these texts have been chosen to be tried in pilot studies with seventh graders. The texts that best discriminate among levels of reading achievement will be used in the final version for large scale testing.

The second example involves a Native American culture without a tradition of written language. The Yupik people of Alaska are concerned about preserving their language, which is rapidly being replaced by English. To assure that new generations have some knowledge of Yupik, members of the community are constructing materials for instruction and reading practice in the Yupik language. They want to write texts with a range of difficulty that can be read by young children and others through the upper grades in school.

It would not make sense in this case either to use or adapt a readability formula. The word and sentence structure of Yupik are quite different from those of English. The same word stem can occur in simple form or with polysyllabic endings, so that word length is not a reliable indicator of difficulty. In English, a sentence like *He made them a large house* has many short words, while in Yupik it would consist of a small number of very long words. Even if a formula could be adapted to take into account these features of word and sentence structure, it would be difficult to check its validity. Since the language has not been written previously, there is no body of written texts that could be used for this purpose.

There are also many practical problems. No one in the Yupik community is trained in education or research. The community has drawn on its members who are fluent in Yupik and sensitive to different styles of speaking used in that language. To construct materials for younger children, they may make use of the style used for telling stories to small children. For older children, they may use the style used when an adult explains information to another adult. Since all of these activities are new and untried, success cannot be guaranteed the first time for every attempt. Something must be done as soon as possible to preserve a language in danger of rapidly being forgotten. By trial and error, members of the Yupik community are finding a reasonable approach to the problem of writing their language at various levels of difficulty. In this way, they are able to construct a written resource to keep the language alive among younger members of the community.

Conclusion

These examples have been taken from contexts in which readability formulas could not have been used, causing researchers and others to try to make the best possible use of available resources. Even in normal circumstances, there are often situations in which the use of formulas would be difficult or inappropriate. People are often at a loss when confronted by such situations and go back to using formulas inappropriately because it is difficult to find alternatives. These examples demonstrate that it is possible to assign difficulty levels to texts without the guidance of formulas.

This chapter has presented a variety of case studies of situations where readability has been estimated by some method other than a readability formula. Science writing requires use of technical and often difficult words to get across concepts, resulting in unrealistically high readability ratings. Trade books may be assigned inaccurate readability ratings because formulas are not sensitive to features (such as literary style) that are important in these books. In countries with languages other than English, it is difficult and time consuming to develop new formulas. When the language has not been written before, time does not permit the use of adapted formulas to grade newly created texts. In all of these situations, a traditional readability formula based on word difficulty and sentence length is either unsuited or unavailable for achieving the desired goals.

These case studies describe some actual situations that are not isolated or unusual. A careful examination can lead to the definition and testing of alternative procedures for assigning grade levels. For as long as there have been readability formulas, there have been teachers and researchers who have had strong reservations about their accuracy and validity. Because of the lack of alternative procedures, the concerns voiced by these critics have had very little effect on the use of formulas. If there are no alternatives, formulas continue to be used. If formulas continue to be used, with all their flaws, it is hard to find alternatives and get them adopted for widespread use. It is hoped that now progress can be made and workable alternatives developed to interrupt this seemingly unbreakable cycle.

References

Davison, A., and Green, G. *Linguistic complexity and text comprehension.* Hillsdale, NJ: Erlbaum, 1988.

Davison, A., and Kantor, R.N. On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly,* 1982, *17,* 187-209.

Gray, W.S., and Leary, B.E. *What makes a book readable?* Chicago: University of Chicago Press, 1935.

Ohanian, S. Ruffles and flourishes. *The Atlantic,* September 1987, *260* (3), 20-22.

Vogel, M., and Washburne, C. An objective method of determining grade placement of children's reading material. *Elementary School Journal,* 1928, *28,* 373-381.

4

Annette T. Rabin

Determining Difficulty Levels of Text Written in Languages Other than English

A discussion of readability measurement in languages other than English requires reviewing the work of researchers both in the United States and abroad. This chapter emphasizes instruments and philosophies representative of researchers outside the United States and relates their research to what has been accomplished here. The chart at the end of this chapter presents a more structured picture of what has been accomplished globally.

The major areas of discussion include:

- A brief history of readability research on foreign language text.
- The validity of using readability instruments to measure texts intended for students studying second languages.
- The development of foreign language readability instruments in the United States.
- The development of readability instruments abroad.
- The use of cloze procedure in the measurement of readability in other countries.
- Some differing philosophies abroad regarding readability evaluation.

The twofold purpose of this chapter is to acquaint readers with the options available for determining the readability of text written in languages other than English and to make readers aware of the techniques available for the development of a readability measurement in any language.

A Brief History of Readability Research in Other Languages

In his review of the use of readability measures for languages other than English, Klare (1974, 1984) pointed out that much of the early research was conducted in the United States for the benefit of English speaking students studying additional languages.

This process began with Tharp in 1939 and has continued. At least seven formulas or their variations for Spanish text have been developed in this country (Crawford, 1984; Garcia, 1977; Gilliam, Peña, & Mountain, 1980; Patterson, 1972; Spaulding, 1951; Thonis, 1976; Vari-Cartier, 1981) as well as instruments for Russian (Rock, 1970), German (Walters, 1966), Hebrew (Nahshon, 1957), Chinese (Yang, 1971), and Vietnamese (Nguyen & Henkin, 1985). Several of the formulas are new, demonstrating an ongoing interest in the evaluation of second language text for language study and for use with recent immigrants.

The earliest readability measures developed in Europe were based on modifications of the Flesch Reading Ease Formula (1950). Kandel and Moles (1958) adapted the instrument to the French language; Fernández Huerta (1959) formulated a Spanish version. De Landsheere (1963, 1970) continued this work, publishing first in French and then in German. Douma (1960) and Brouwer (1963) used variations of the Flesch formula in the development of measures for text in the Dutch language.

Intensive research on more original instruments began in Europe in the early 1960s. In Sweden, Björnsson (1968a, 1968b, 1983) abandoned the regression equation in favor of an additive formula. This technique was later elaborated by Bamberger and Vanecek (1982, 1984) working on German text in Austria and simplified for use with English materials by Anderson (1983) in Australia.

The first original German formula developed in Europe was that of Fucks (1955). Sentence length was multiplied by word length to yield a difficulty level. This instrument produced results similar to those from the Fry graph but was judged inappropriate for the German language, probably due to longer words in German than in English. Other European research in German followed (Bamberger, 1973; Briest, 1974, Dickes & Steiwer, 1977; Groeben, 1972; Nestler, 1977).

In Holland the research of van Hauwermeiren (1972) resulted in six new formulas, each using a different combination of variables. Two additional investigations followed (Zondervan, van Steen, & Gunneweg, 1976; Staphorsius & Krom, 1985).

Meanwhile, Henry (1975) set out to develop a readability instrument especially for French. Two studies had already been done at the University of Liège. Foucart (1963) had shown that the difficulty levels of popular texts were lower than those of texts rejected. De Landsheere (1964) found that, in some cases, a high human interest score on the Flesch could provoke a rejection of, instead of an attraction to, the material.

Henry felt this reaction raised the question of whether a readability instrument developed on one certain group of books or subject matter could be applied to evaluate other materials.

Richaudeau (1973, 1981, 1985) undertook research which convinced him that the simplest sentence is not necessarily the most easily understood. He, like Kintsch (1979), conducted investigations which demonstrated that certain transformations (whether long or short) appear to stay with the reader longer. Richaudeau proposed an experimental formula that spoke to the syntactic complexity of the reading material rather than its grade level.

Major research in the development of instruments for use with Spanish text has emanated from Venezuela (Gutiérrez et al., 1972; Morles, 1975, 1981; Rodríguez Trujillo, 1978, 1980, 1983) and Spain (López Rodríguez 1981, 1982; Rodríguez Diéguez, 1983, 1987). Several of the Venezuelan projects were initiated under John Bormuth at the University of Chicago. Original investigation in Spain did not begin until the 1980s. Research has been ongoing in both countries.

Although English is the basic language used in the United Kingdom, a survey of readability in other countries would be incomplete without mention of research conducted there. Two major works are those of Gilliland (1972, 1975) and Harrison (1980). Gilliland's is a more theoretical survey of variables–linguistic, phonological, and physical. Harrison's research concerns the practical use of readability measurement in the classroom. It includes results of two government studies related to the suitability of the textbooks used by British school children: the Bullock committee inquiry into the teaching of English (Department of Education and Science, 1975) and the Effective Use of Reading Project (Lunzer & Gardner, 1979).

The Bullock Report stresses the importance of assessing difficulty levels of texts to match children to their study materials. The result has been increased attention in the United Kingdom to the use of readability measurements, with instruction in their use included in many pre and inservice teacher education programs. In a survey conducted as part of the Lunzer and Gardner project, computerized versions of six readability formulas were used to evaluate 125 texts from four subject areas. Results were compared with teacher judgment. The Dale-Chall formula, though time consuming to calculate, was found to be the best overall. Scores yielded by SMOG and FOG were higher than teacher estimates. On difficult material, scores at the upper levels were hard to interpret.

The Validity of Using Readability Instruments to Measure Texts

In this monograph and elsewhere, Klare (1974, 1984) has discussed the choice of criteria and variables on which to base readability instruments. Of the more than 250 variables studied, word length (alone or in combination with word frequency) and sentence length account for most of the variance in the measurement of reading materials.

Laroche (1979) considered the use of the cloze procedure or word frequency lists questionable as criteria for establishing formulas. A cloze test (Taylor, 1953) is often administered to a criterion population and then used as the basis for a readability instrument.

Laroche noted that results of a cloze test are said to reflect a basic intuition about the structure and vocabulary of the target language. He saw no such population available in the case of foreign language reading materials. Likewise, he considered the use of word frequency lists a fallacy, pointing out that the use of a word is related to the "language bath" in which one is immersed, and the linguistic ambiance of the native speaker would differ from that of the language student.

In discussing the psychosocial dimensions of language acquisition, Ervin-Tripp (1973) observed that the age at which a second language is learned affects the body of meaning acquired. While a child's thinking might be more oriented toward personal needs, an adult's speech would reflect a more complex development of knowledge and skills. This would influence vocabulary usage. There is also a difference between the way one mentally processes the language (or languages) with which one has grown up and those newly acquired. This would support Laroche's contention that there would not be the same (or even similar) intuition for language among those in a nonnative criterion population as in a native one.

Laroche appealed for greater consideration of linguistic variables and, citing Bormuth's work (1970), greater collaboration between the disciplines in the study of reading comprehension. Recognizing the need for readability measurement in foreign language instruction, he recommended an instrument that would take into account cognate count and frequency, sentence length as a reflection of syntactic complexity, and phrase structure complexity.

In contrast to Laroche and Ervin-Tripp, Schulz (1981) claimed that psycholinguists assume that once sound-symbol correspondences have been established and students are familiar with a body of vocabulary and major patterns of syntax and morphology, reading in a second language is identical to reading in one's native tongue. Schulz quoted no source for this theory, however, and described a study by Clarke (1980) that seemed to refute this stand. Clarke administered an ESL test to two groups of Spanish speaking adults who had been grouped into good and poor readers in their native language. Results showed much less variation in their scores on the ESL test than on a Spanish reading test, demonstrating the leveling effect of the belated acquisition of the second language.

Schulz and Laroche were concerned that there be a way of avoiding frustrational reading of literary texts in a foreign language. Though aware of Laroche's position, Schulz chose to dismiss it on the grounds that the limited research available supported the use of similar linguistic criteria for measuring readability in all western languages. She dealt only minimally with the question of the appropriateness of current evaluative instruments for nonnative speakers. In discussing the use of cloze procedure, she observed that a foreign language student might guess the meaning of a missing word from context and not be able to supply the specific foreign word.

Foreign Language Readability Instruments in the United States

Whether because of the popularity of Spanish language study in our schools or the proximity of Spanish speaking countries and the resultant immigration of their citizens, several instruments for the evaluation of Spanish text have been developed in the United States.

Following the Dale-Chall model (1948), Spaulding (1951, 1956) developed a formula using the two variables of word usage or frequency as measured by a density calculation and sentence complexity as measured by average sentence length. Word frequency was based on the number of words in a passage that did not appear on the Buchanan list (1941) of the 1,500 most frequently used words in Spanish. Spaulding's formula was adopted by inter-American groups.

Spaulding's procedure was later adapted by Patterson (1972) for use by religious workers dealing with readers with minimal reading ability. Thonis (1976) used Patterson's descriptions of the reading skills needed to understand materials in the various categories in Spaulding's formula to establish grade levels. Since the procedures followed are not clear, great credence has not been given to Thonis' research.

As with English language measures, faster, simpler methods of computing the readability of Spanish language materials were

64

sought. At least four studies (Crawford, 1984; Garcia, 1977; Gilliam, Peña, & Mountain, 1980; Vari-Cartier, 1981) were based on the Fry Readability Graph (1968, 1977), which uses word and sentence lengths as variables.

In all four projects, it was determined that the syllable count was significantly higher for a 100 word passage in Spanish than for a similar one in English, probably due to the fact that all vowels are pronounced in Spanish.

Gilliam, Peña, and Mountain evaluated twenty-two books designed for grades one through three using Fry's original graph. They concluded that it would be necessary to subtract 67 from the average number of syllables for the closest equivalencies between Spanish and English. Both Garcia (1977) and Vari-Cartier (1981) determined that sentence length also should be adjusted. Garcia's criterion was a basal reading series in Spanish. Vari-Cartier used 127 samples of Spanish prose materials to develop the FRASE (Fry Readability Adaptation for Spanish Evaluation) graph. She suggested that the procedures used in developing this graph could be applied to languages other than Spanish by adjusting the parameters for minimum and maximum sentence and syllable counts and readability designations.

Crawford's research (1984) was supported by the U.S. Department of Education under the Bilingual Education Act. He chose as his criterion the Laidlaw series of elementary Spanish texts (Pastor et al., 1971) after determining that the progression of increase for average sentence length and number of syllables per 100 words was more regular in this series than in the nine other series he evaluated.

After exhaustive international correspondence, Schwartz (1975) concluded that no adequate measure existed for instructional materials at the elementary level in German and she adapted the Fry graph to that language. Using samples from a series of West German basal readers dating from the post World War II era as the criterion, she determined that the longer length of German words results in a count of from 25 to 37 syllables higher than in English. The number of sentences per 100 words, however, was very close for corresponding grade levels.

Those developing other formulas in the United States for use with foreign text also found semantic (word length or frequency) and syntactic (sentence length) factors to be the most predictive, though in the case of Oriental languages a need existed for additional considerations. Yang (1971) included character factors in his variables. A Vietnamese instrument (Nguyen & Henkin, 1982) included tonal marks, word marks, and hyphens as part of letter count.

Development of Readability Instruments Abroad

Although research in the United States has dictated the use of a limited number of linguistic variables, instruments developed abroad have sometimes been more complex.

French. In France, Henry (1975) developed three formulas, an eight variable instrument (he considered it the most valid) that was very complicated and required a knowledge of linguistics; a computerized formula with limited practicality; and a formula designed for manual use by teachers. This last formula took into account only three variables: number of words per sentence, number of words absent from the Gougenheim et al. word list (1967), and first names only used with exclamation points and quotation marks.

All three instruments can be used on three levels: grades five and six, eight and nine, and eleven and twelve, allowing for the evaluation of the same text at each grade level. Measurement is in terms of percentage on a cloze, with 35-45 percent indicated as the comfort zone. Anything below is too difficult; anything above is not sufficiently challenging. Graphs eliminate the need for calculations.

Spanish. Gutiérrez et al. (1972), working in Venezuela, were responsible for what appears to be the first original readability formula for use with Spanish text developed outside the United States. A multiple regression equation with cloze as the criterion, it was validated only at the sixth grade level. This research was conducted under Venezuela's Ministry of Education in response to the great need for a method of matching students and their text materials. Unfortunately, many of Gutiérrez's compatriots neither understood nor were ready to accept the concept of readability measurement, and the procedure was never widely used.

Since publishers' evaluations of the readability levels of Venezuelan textbooks are still inadequate, and many teachers have only the equivalent of a high school education, the need persists for some type of readability instrument. Currently, Rodríguez Trujillo (undated) is attempting to develop an evaluative technique that can be used to determine both the difficulty levels of educational materials and the reading ability of the students using them. A procedure for Spanish modeled on Carver's Rauding Scale (1976) is being considered.

In Spain, López Rodríguez (1981, 1983) studied twenty-six linguistic variables, selecting seven of these for her first formula. Among those used was a list of common vocabulary by García Hoz (1953). Her criterion was derived from cloze tests, each administered to ten students. Rodríguez Diéguez (1983) added eight variables to his predecessor's list, using twelve in his instrument. His criterion was developed from 123 cloze tests, each also administered to ten subjects. Currently, he is working on a formula that will extend to the end of college in two year intervals.

Swedish. Many of the aforementioned readability formulas are in the form of regression equations. Björnsson (1968a, 1968b) of Sweden was a pioneer in the development of additive formulas, a technique in which linguistic factors are simply added together and the result compared with a predetermined set of criteria.

This was not done arbitrarily. Björnsson worked in several languages. In one of his many research studies (1974) he used 100 texts. Their levels of difficulty were judged by two groups of judges, each evaluating half of the same texts. Correlation was quite high between the average assessments of the two groups. Based on his results he concluded that, contrary to traditional belief, judges' ratings would be reliable if three conditions existed: (1) they were made by a sufficiently large group of persons, (2) the passages were relatively long, and (3) the range of difficulty in the text battery was wide. The average correlation coefficient for groups of six judges was .94, as opposed to an average of .99 for twenty-four or more. Björnsson originally attempted to develop a regression equation for his Lix (short for *läsbarhetsindex*) readability index in Swedish. Using all 100 texts, he derived an acceptable equation based on the calculation of multiple correlation. However, when he divided his texts in half and recalculated the equation, he found he would have obtained completely different equations and coefficients of validity if his study had happened to include the first fifty texts or the second separately. Björnsson concluded that regression equations were closely dependent on the composition of the criterion and not suitable as readability formulas and so he turned to the additive method (Björnsson, 1983). After experimenting in Swedish with twelve variables, he settled on two as the best predictors of text difficulty–sentence length and percentage of long words (in this case those with seven or more letters).

German. Variables used in German readability measures have been varied and plentiful. Dickes and Steiwer (1977) developed several formulas, one with as many as eight variables. Groeben (1972) used the level of abstractness of words to determine text difficulty. Nestler's formula (1977) dealt with the conceptual levels of words in three categories: (1) generally known words, (2) hard words, and (3) rare professional words.

Bamberger (1973), working alone and eventually with others (Bamberger & Vanecek, 1982, 1984; Bamberger & Rabin, 1984), initiated a project to measure German textual materials by both subjective and objective means. A "readability profile" composed of five nonlinguistic variables–content, organization, print, style, and motivation–was used in combination with a series of regression formulas.

The checklist of more than thirty items yielded grade levels that could be compared with those given by the formulas. When the combination of the language difficulty and the readability profile was applied to several hundred books in a cross validation, it was demonstrated that in approximately 70 percent of the cases, the grade level yielded by the profile was similar to that resulting from the use of the formulas. They felt this was an indication of the usefulness of readability formulas.

Learning of Björnsson's additive formula, Bamberger and Vanecek elaborated on the technique by adding other linguistic factors. They used as their criteria 120 children's storybooks and 200 nonfiction textbooks that previously had been arranged into grade levels through the use of pooled assessment and the application of readability formulas. Tables were developed that showed the average values of six linguistic factors, plus the calculation of Lix, by grade level, for works of fiction and nonfiction in the German language.

Both these tables and the readability profile were designed so that educators could discern which individual variable or combination might be causing difficulty. These variables then could be handled instructionally. Much of the Austrian evaluative procedure has now been computerized.

Danish. Denmark also has benefited from Björnsson's research on Lix (Jansen, 1987). In the 1960s, newspaper publishers became interested in widening the use of newspapers in schools. They contacted the Danish Institute for Educational Research to obtain the level of linguistic difficulty of a number of daily papers. Research already had been started by Jesper Florander and Mogens Jansen (1966) when a query to Swedish colleagues brought news of Björnsson's studies (1964). Since Swedish and Danish are similar languages, the Danes opted to adapt Lix to their purposes. The current Danish readability evaluation represents the sum of the average length of meaning (sentence length) and the percentage of long words (words of more than six letters).

Danish researchers see readability measurement as the interaction among three components: linguistic, represented by Lix; visual, including typography, layout, paper, and print; and the "contents of the text," defined as the personal interest to the reader of the contents of the text. These components relate to three levels of readers in a 2x3 schema: rebus, those who are either beginners or disabled readers; transition, those having reached a degree of reading competency; and content, those able to choose texts solely on the basis of content without concern for external appearance or language. All Danish teaching materials, all children's books, and many books for young people and adults have been evaluated since 1970. Through the ongoing use of Lix, it has been possible to follow the development of the linguistic levels of books. Nonfiction works for nine to thirteen year olds have become more difficult since the late 1960s, and most children have difficulty reading many of the nonfiction books published for them. The Lix Committee has published three official reports on their efforts (Jakobsen, 1971, 1976, 1983).

English. Anderson (1967, 1971), who had previously experimented with the use of cloze procedure and a readability scale to evaluate the readability levels of children's books used in Australian schools also became interested in Lix. He developed conversion tables for expressing Lix scores in grade levels for English language materials. As an outcome of his calculations, Anderson (1983) noticed that readability estimates could be obtained by simply calculating the average number (or rate) of long words per sentence. He called his new measure Rix (rate index). Actually, sentence length is still involved indirectly, as the number of sentences and the number of long words must be counted and divided.

Use of Cloze Procedures in Foreign Languages

There is controversy regarding the use of cloze procedure in determining the readability of written materials. This controversy is based on the fact that cloze is a subjective evaluation that mirrors the language ability and background of information of the person taking the test. Also, some researchers feel that multiple cloze passages should be developed from each piece of material for the results to be valid. For example, a test deleting every fifth word should be prepared in five versions, omitting a different word each time. Though these views are shared by other countries, for want of a better technique, cloze procedure is widely used.

A good example is the extensive research on the use of cloze procedure in the measurement of the readability of Spanish language reading materials by researchers in Venezuela (Bastidas, Calderón, & Bravo, 1981; Morles, 1975, 1981; Rodríguez Trujillo, 1978, 1980, 1983) and Spain (López Rodríguez, 1983). Using Bormuth's levels (1971), Morles (1981) discovered that in addition to being appropriate for the determination of the student's ability to comprehend a text, cloze could be used as an indication of what percentage of the total group could handle the material by determining how many students had scored more than 58 percent. Bastidas, Calderón, and Bravo (1981) found that when cloze tests contained at least fifty items, there was a high correlation between parallel forms from the same material.

The latter observation was in contrast to what Derakshani (1980) learned when he used cloze passages to determine the readability level of a Persian text on volleyball destined for the popular market in Iran. Derakshani compared scores on five parallel cloze tests with those on a twenty-five item language achievement and a ten item multiple choice comprehension test. Unlike Bastidas, he found that there was not always the same mean score for different versions of a particular passage. He was able to demonstrate statistically that there was a positive relationship between skill in the use of the target language and achievement on a cloze passage. These findings are similar to those of Entin (Entin, 1980; Entin & Klare, 1985) in the United States. Working with cloze tests based on two of the five possible versions, she found that in sixteen of twenty-four comparisons there were significant differences at the .05 level.

Mikk, Sepp, and Hanson (1973) investigated the possibility of using cloze procedure to evaluate the readability of Estonian text. Research in which subjects were presented with a progressively larger number of words on either side of a deleted one, starting with three words, led to the conclusion that it was preferable to omit every seventh word instead of every fifth. Two of their experiments indicated that requiring the exact words in the blanks yielded a better indication of the pupils' achievement and mental abilities. Accepting alternative answers as long as they fit the content was a better indicator of the difficulty of the text. They concluded, however, that it was more efficient to count only the exact word correct and determined that cloze tests consisting of about fourteen pages of a book were needed for accurate evaluation, with three pages recommended if all of the words fitting the content were considered. Since words were deleted with less frequency, it was felt that the percentages for the various instructional levels usually used were not appropriate, and further research was indicated.

Sukeyori (1957) tested the applicability of cloze procedure to the Japanese language. Experiments were conducted to determine what percentage of material should be deleted and whether it was preferable to delete letters or words. Results showed that a deletion pattern of 10 to 20 percent of the words was more suitable than deletion of letters.

Other languages in which cloze has been used for readability measurement include French (Henry, 1973; De Landsheere, 1972, 1973), Korean (Taylor, 1956), and Vietnamese (Klare, Sinaiko, & Stolurow, 1971). Klare (1974, pp. 95-96) offers additional references for cloze research in some of the above languages, as well as for Thai and for foreign speakers of English in Papua New Guinea.

Differing Philosophies

As with Kintsch (1979) in the United States, there are those abroad who would take other variables into consideration in determining the difficulty of reading materials.

Among them is Richaudeau (1973, 1981), whose research convinced him that certain transformations, whether long or short, appear to stay with the reader longer. He criticized the validation of readability formulas with cloze procedure, pointing out that greater ability to complete a cloze test is directly related to the redundant material measured by such formulas. Richaudeau did not advocate abandoning formulas altogether, but he felt teachers and publishers should remember this and realize that the more redundancy there is in a text, the less interest it holds.

Richaudeau stressed the importance of anticipation over meaning, observing that we have built a network of neurological pathways over which our knowledge of certain concepts travel. Stimuli from outside reactivate these concepts. He rejected the complete sentence as a unit, maintaining that the sousphrase – probably what we would call the clause – punctuated with a period, semicolon, colon, or dash is the important unit.

Important aids to the readability of such clauses and sentences include:

- The placement at the beginning of a clause of important words such as the subject, verb, and principal adjective.
- The use of short, common words.
- The use of anticipated words as cues, e.g., I have ____ some ____ which ____.

- The use of an affirmative formula at the beginning, e.g., This is why _____.
- The limitation of number of words separating subject and verb.

Richaudeau's experimental formula measured the number of words retained after reading a sentence or clause based on three variables: the relationship between word length and sentence length as plotted on a graph, whether the sentence began with an affirmative formula, and the number of words between the subject and the verb.

Platzack (1974) in Sweden conducted readability experiments that included research on how physical, syntactic, semantic, and contextual cues influence the difficulty level of written materials. He reached the following conclusions:

- Physical cues, like punctuation marks and short structure words, help to set off decoding units.
- A text in which relative pronouns have been deleted is often less readable than when these are present.
- Eye-voice span becomes wider when certain short words are present as cues to underlying structure.
- A sentence in which an adverbial clause is placed between the verb and the object of the main clause is more difficult to read than one in which the adverbial clause is placed after the object.

Platzack maintained that a sentence with a mean length of approximately thirteen words is easier to read than one in which the mean sentence length is less than nine words, assuming "the texts are of the same difficulty." Quoting Smith and Holmes (1973), he observed that long term memory can take in new materials every third to fifth second only and deduced that someone who reads 200 words per minute would be able to read 10 to 17 words during the interval when long term memory is locked. A short sentence, therefore, might not make it into long term memory. Readers would either have to wait or to read more than one sentence before they were able to store what had been read.

Conclusion

The development of readability instruments for the evaluation of foreign language text has been an ongoing process. It began with research in the United States on materials used in second language learning and has extended to most of the civilized world, first by the adaptation of formulas intended for English language text and then by original investigation.

Although research in the United States has tended to narrow the selection of variables contributing to text difficulty to semantic (word length) and syntactic (sentence length) factors, there are foreign researchers who have seen fit to increase these. A notable exception is Björnsson of Sweden, whose investigations produced the additive formula Lix.

Both subjective judgment and cloze procedure have been used extensively abroad in the development of criteria on which to base readability instruments. In situations where formulas have not been feasible, cloze procedure (often using Bormuth's levels) has been adapted to local needs.

There are those who feel that the variables used in the development of readability measures for second language text should differ from those variables used for one's native tongue. Concern has been voiced by investigators abroad who believe, as some American researchers do, that such factors as physical, psycholinguistic, and contextual cues should be considered when evaluating the difficulty level of written text.

Finally, there appears to be a place for continued investigation into the factors that affect the comprehensibilty, or difficulty level, of textual materials in both foreign languages and English. It is apparent, as one studies extant research, that the procedures needed for developing a readability measure for any language are readily available to anyone interested in going through the process.

References

Anderson, J. *A scale to measure the reading difficulty of children's books.* St. Lucia, Australia: University of Queensland, Faculty of Education, 1967.

Anderson, J. Lix and Rix: Variations on a little known readability index. *Journal of Reading,* 1983, *26,* 490-496.

Anderson, J. Selecting a suitable reader. Procedures for teachers to assess language difficulty. RELC *Journal*, 1971, *2*, 35-42.

Bamberger, R. *Lese-Erziehung*. Vienna, Austria: Jugend und Volk, 1973.

Bamberger, R., and Rabin, A.T. New approaches to readability: Austrian research. *The Reading Teacher*, 1984, *37*, 512-519.

Bamberger, R., and Vanecek, E. *Die Lesbarkeit oder die Schwierigkeitsstufen von Texten in deutscher Sprache*. Unpublished research study, Vienna, Austria, 1982.

Bamberger, R., and Vanecek, E. *Lesen-Verstehen-Lernen-Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Vienna, Austria: Jugend und Volk, 1984.

Bastidas, E., Calderón, I., and Bravo, Ch. *Confiabilidad del procedimiento "cloze."* Unpublished master's thesis, Central University of Venezuela, 1981.

Bhagoliwal, B.S. Readability formulae: Their reliability, validity, and applicability in Hindi. *Journal of Education and Psychology*, 1961, *19*, 13-26.

Björnsson, C.H. *Läsbarhet*. Stockholm, Sweden: Bokforlaget Liber, 1968a.

Björnsson, C.H. *Läsbarhet*. Stockholm, Sweden: Skolöverstyrelsen, 1964.

Björnsson, C.H. *Läsbarhetsprövning av engelsk skoltext*. Technical Report No. 53. Stockholm, Sweden: Pedagogical Center, 1974.

Björnsson, C.H. *Lesbarkeit durch Lix*. Technical Report No. 61. Stockholm, Sweden: Pedagogical Center, 1968b.

Björnsson, C.H. Readability of newspapers in 11 languages. *Reading Research Quarterly*, 1983, *18*, 480-497.

Bormuth, J.R. *Development of standards of readability: Toward a rational criterion of passage performance*. U.S. Department of Education Project No. 7-0052. Chicago: University of Chicago Press, 1971.

Bormuth, J.R. *On the theory of achievement test items*. Chicago: University of Chicago Press, 1970.

Briest, W. Kann man Verständlichkeit messen? *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 1974, *6*, 543-563.

Brouwer, R.H. Onderzoek naar de Leesmoeilijkheid van Nederlands Proza. *Paedagogische Studiën*, 1963, *40*, 454-464.

Buchanan, M.A. A graded Spanish word book. Toronto, Ontario: University of Toronto Press, 1941.

Carver, R.P. Measuring prose difficulty using the Rauding scale. *Reading Research Quarterly*, 1976, *11*, 660-685.

Clarke, M.A. The short circuit hypothesis of ESL reading. *Modern Language Journal*, 1980, *64*, 203-209.

Crawford, A.N. A Spanish language Fry-type readability procedure: Elementary level. *Bilingual Education Paper Series*, *7*. Los Angeles, CA: Evaluation, Dissemination, and Assessment Center, California State University, 1984.

Dale, E., and Chall, J.S. A formula for predicting readability. *Educational Research Bulletin*, 1948, *27*, 11-20, 37-54.

De Landsheere, G. Application du cloze test de W.L. Taylor à la langue française. *Scientia Paedagogica Experimentalis*, 1972, *9*, 207-256.

De Landsheere, G. *Einfürung in die pädagogische Forschung*, second edition. Weinheim: Beltz, 1970, 225-234.

De Landsheere, G. *Le test de closure: Mésure de la lisibilité et de la compréhension*. Paris, France: Nathan; Brussels, Belgium: Labor, 1973.

De Landsheere, G. Pour une application des tests de lisibilité de Flesch à la langue française. *Le Travail Humain*, 1963, *26*, 141-154.

De Landsheere, G. *Recherche sur l'évaluation objective de la lisibilité des manuels scholaires et des tests*. Association Internationale de Pédagogie Expérimentale de Langue Française. Eleventh International Colloquium, Liège, 1964, 73-97.

De Landsheere, G. *Lecteures et lectures, recherches sur l'évaluation et le contrôle objectif*. Association Internationale de Pédagogie Expérimentale de Langue Française. Thirteenth International Colloquium, Geneva, April 1966.

Department of Education and Science. *A language for life, the Bullock report.* London: Her Majesty's Stationery Office, 1975.

Derakshani, M. Cloze procedure as a measure of readability of Persian texts. *Psychologia: International Journal of Psychology in the Orient*, 1980, *20*, 193-194.

Dickes, P., and Steiwer, L. Ausarbeitung von Lesbarkeitsformeln für die deutsche Sprache. *Zeitschrift für Entwicklungspsychologie und padagogische Psychologie*, 1977, *9*, 20-28.

Douma, W.H. De Leesbaarheid van Landbowbladen: een Onderzoek naar en een Toepassing van Leesbaarheidsformules. Bulletin No. 17. *Sociologie en Sociographie.* Netherlands: Landbouwhogeschool te Wageningen, 1960.

Entin, E.B. *Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages.* Unpublished doctoral dissertation, Ohio University, 1980.

Entin, E.B., and Klare, G.R. Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages. In B. Hutson (Ed.), *Advances in reading/language research*, volume 3. Greenwich, CT: JAI Press, 1985.

Ervin-Tripp, S.M. *Language acquisition and communicative choice.* Stanford, CA: Stanford University Press, 1973.

Farr, J.N., Jenkins, J.J., and Paterson, D.G. Simplification of Flesch reading ease formula. *Journal of Applied Psychology*, 1951, *32*, 221-233.

Fernández Huerta, J. Medidas sencillas de lecturabilidad, *Consigna*, 1959, *214*, 29-32.

Flesch, R.F. A new readability yardstick. *Journal of Applied Psychology*, 1948, *32*, 221-233.

Florander, J., and Jansen, M. Om letlaeseligheden af nogle dagblade – specielt af Berlingske Tidende (About the readability of some daily papers – especially of Berlingske Tidende). Copenhagen, 1966.

Foucart, C. *Etude de textes d'un livre de lecture du degré supérieur primaire. Essai de détermination des caractères qui influencent le choix et le rejet des lectures.* Unpublished memorandum, University of Liège, France, 1963.

Fry, E.B. A readability formula that saves time. *Journal of Reading*, 1968, *2*, 513-516, 575-578.

Fry, E.B. Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading*, 1977, *21*, 242-252.

Fucks, W. Unterschiede des Prosastils von Dichtern und anderen Schriftstellern. *Sprachforum, 3/4.* Münster: Köln, 1955.

Garcia, W.F. Assessing readability for Spanish as a second language: The Fry graph and cloze procedure. *Dissertation Abstracts*, 1977, *38*, 136a.

García Hoz, V. Vocabulario usual, vocabulario común y vocabulario fundamental. Determinación y análisis de sus factores. Madrid, Spain: CSIC, 1953.

Gilliam, B., Peña, S.C., and Mountain, L. The Fry graph applied to Spanish readability. *The Reading Teacher*, 1980, *3*, 426-430.

Gilliland, J. *Readability.* London: University of London Press, 1972; Hodder and Stoughton, 1975.

Gougenheim G., Rivenc, P., Michea, R., and Sauvageot, A. L'élaboration du français fondamental. Paris, France: Didier, 1967, 69-113.

Groeben, N. *Die Verständlichkeit von Unterrichtstexten, Dimensionen Kriterien rezeptiver Lernstadien.* Münster: Verlag Aschendorff, 1972.

Gunning, R. *The technique of clear writing.* New York: McGraw-Hill, 1952; revised 1968.

Gutiérrez de Polini, L.E. et al. *Investigación sobre lectura en Venezuela.* Paper presented before the Primeras Jornadas de Educación Primaria. Caracas, Venezuela: Ministry of Education, 1972.

Harrison, C. *Readability in the classroom.* Cambridge: Cambridge University Press, 1980.

Henry, G. *Comment mésurer la lisibilité.* Brussels, Belgium: Editions Labor; Paris, France: Fernand Nathan, 1975.

Henry, G. *Une technique de mésure de la lisibilité specifique de la langue française.* Unpublished doctoral dissertation, University of Liège, France, 1973.

Jakobsen, G. (Ed.). Dansk Lix 70. *Laesepaedagogen*, 1971.

Jakobsen, G. (Ed.). Dansk Lix 75. *Laesepaedagogen*, 1976.

Jakobsen, G. (Ed.). Dansk Lix 83. *Laesepaedagogen*, 1983.

Jansen, M. *Danish Lix: A Danish readability formula.* Unpublished paper, Danish National Association of Reading Teachers, 1987.

Johnson, G.R. An objective method of determining reading difficulty. *Journal of Educational Research*, 1930, *21*, 283-287.

Kandel, L., and Moles, A. Application de l'indice de Flesch à la langue française. *Cahiers d'Etudes de Radio-Télévision*, 1958, *19*, 253-274.

Kintsch, W. On modeling comprehension. *Educational Psychologist*, 1979, *19*, 253-274.

Klare, G.R. Assessing readability. *Reading Research Quarterly*, 1974, *10*, 62-102.

Klare, G.R. Readability. In P.D. Pearson (Ed.), *Handbook of reading research.* New York: Longman, 1984, 681-744.

Klare, G.R., Sinaiko, H.W., and Stolurow, L.M. *The cloze procedure: A convenient readability test for training materials and translations.* Arlington, VA: Institute for Defense Analysis, 1971.

Laroche, J.M. Readability measurement for foreign language materials. *System*, 1979, *7*, 131-135.

López Rodríguez, N. *Como valorar textos escolares.* Madrid, Spain: Cincel, 1982.

López Rodríguez, N. *Fórmulas de legibilidad para le lengua castellana.* Unpublished doctoral dissertation, University of Valencia, Spain, 1981.

López Rodríguez, N. Una técnica para medir la comprensión lectora: El test cloze. *Enseñanza*, 1983, *1*.

Lunzer, E.A., and Gardner, W.K. (Eds.). *The effective use of reading.* London: Heinemann, 1979.

Mikk, J., Sepp, M., and Hanson, M. How to use cloze readability procedure. Copy of a summary of a Russian article in *Nöukogude Kool*, 1973, *20*, 32-34.

Morles, A. *The scoring of cloze comprehension tests in the Spanish language.* Unpublished master's thesis, University of Chicago, 1975.

Morles, A. Mendición de la comprensibilidad de materiales escritos mediante pruebas "cloze." *Lectura y Vida*, 1981, *2*, 16-18.

Nahshon, S. *Readability measurement of Hebrew prose.* Unpublished doctoral dissertation, Teachers College, Columbia University, 1957.

Nestler, K. Zur Ermittlung von Voraussetzungen für die Formulierung von Normen für die angemessene sprachliche Gestaltung von Lehrtexten. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikation forschung*, 1977, *1*.

Nguyen, L.T., and Henkin, A.B. A readability formula for Vietnamese. *Journal of Reading*, 1985, *26*, 243-251.

Park, Y. An analysis of some structural variables of the Korean language and the development of a readability formula for Korean textbooks. *Dissertation Abstracts*, 1974, *35*, 946A.

Pastor, A., Guzmán Vda de Capo, R., Gómez Tejera, C., and Hester, K.B. *Por el mundo del cuento y la aventura.* River Forest, IL: Laidlaw Brothers, 1971.

Patterson, F.W. *Como escribir para ser entendido.* El Paso, TX: Casa Bautista de Publicaciones, 1972.

Platzack, C. *Språket och läsbarheten* (with English summary). Lund, Sweden: Gleerup, 1974.

Richaudeau, F. *Le langage efficace.* Paris, France: Retz, 1973.

Richaudeau, F. Some French work on prose readability and syntax. *Journal of Reading*, 1981, *24*, 503-508.

Richaudeau, F. Translated by Yvonne V. Basista. The reading process in six diagrams. *Journal of Reading*, 1985, *28*, 504-512.

Rock, E.L. A readability graph for Russian. *Dissertation Abstracts*, 1970, *31*, 567-A.

Rodríguez Diéguez, J.L. Evaluación de textos escolares. *Revista de Investigación Educativa*, 1983, *1*, 259-279.

Rodríguez Diéguez, J.L. *Predicción de la lecturabilidad de textos en castellano: Una propuesta y algunas sugerencias.* Paper presented at the Fifth European Reading Conference in Salamanca, Spain, 1987.

Rodríguez Trujillo, N. An adaptation of the cloze procedure to the Spanish language. *Dissertation Abstracts*, 1978, *38*, 7191A.

Rodríguez Trujillo, N. Determinación de la comprensibilidad de materiales de lectura por medio de variables lingüísticas. *Lectura y Vida*, 1980, *1*, 29-32.

Rodríguez Trujillo, N. El procedimiento cloze: un procedimiento para evaluar la comprensión de lectura y la complejidad de materiales. *Lectura y Vida*, 1983, *4*, 4-13.

Rodríguez Trujillo, N. *Establecimiento de una escala para determinar la comprensibilidad de materiales de lectura y la comprensión de lectura de grupos de alumnos.* Pamphlet. Banco del Libro, Programa de Bibliotecas Escolares, Caracas, Venezuela, undated.

Schulz, R.A. Literature and readability: Bridging the gap in foreign language reading. *Modern Language Journal*, 1981, *65*, 43-53.

Schwartz, R.E.W. *An exploratory effort to design a readability graph for German material.* Unpublished study, State University of New York at Albany, 1975. (ED 123 565)

Smith, F., and Holmes, D.L. The independence of letter, word, and meaning identification in reading. In F. Smith (Ed.), *Psycholinguistics and reading.* New York: Holt, Rinehart and Winston, 1973, 50-69.

Spaulding, S. A Spanish readability formula. *Modern Language Journal*, 1956, *40*, 433-441.

Spaulding, S. Two formulas for estimating the reading difficulty of Spanish. *Educational Research Bulletin*, 1951, *30*, 117-124.

Staphorsius, G., and Krom, R.S.H. *CITO leesbaarheidsindex voor het basisonderwijs.* Arnheim: Central Institute voor Toets Ontwikkeling, 1985.

Sukeyori, S. A study of readability: Applications of cloze procedure to Japanese language. *Japanese Journal of Psychology*, 1957, *28* (English summary).

Taylor, W.L. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, *30*, 415-433.

Taylor, W.L. Recent developments in the use of cloze procedure. *Journalism Quarterly*, 1956, *33*, 42-48, 99.

Tharp, J.B. The measurement of vocabulary difficulty. *Modern Language Journal*, 1939, *24*, 169-178.

Thonis, E.W. *Literacy for America's Spanish speaking children.* Newark, DE: International Reading Association, 1976.

van Hauwermeiren, P. *De Lesbaarheidsmeting, Topasselijkheid op het Nederlands.* Unpublished doctoral dissertation, University of Louvain, Belgium, 1972.

Vari-Cartier, P. Development and validation of a new instrument to assess the readability of Spanish prose. *Modern Language Journal*, 1981, *65*, 141-148.

Walters, T.W. Readability in German: Some levels of difficulty encountered in reading professional theological literature in German. *Dissertation Abstracts*, 1966, *27*, 2520A.

Yang, S.J. A readability formula for Chinese language. *Dissertation Abstracts*, 1971, 6536A.

Zondervan, T., van Steen, P., and Gunneweg, G. De Leesbaarheid van Basisschoolteksten: Objectieve Ordeningscriteria voor instructieve teksten. *De Nieuwe Taalgids*, 1976, *62*, 426-445.

Representative Readability Formulas and Research in Languages Other than English*

Author and Date of Publication	Formula or Research	Observations
Chinese		
Yang, 1971	Y = 14.95961 (CONSTANT) + 39.077461 × (WORDLIST) − 2.4849 × (STROKES) + 1.11506 × (FULLSEN) FULLSEN = proportion of words in 5,600 simple word list STROKES = average number of strokes per character	Criteria – results of standardized tests based on 85 passages from modern Chinese writings administered to first and second year Taiwanese high school students. Multiple correlation of .80 with independent (character, word, and sentence factors) and dependent (comprehension) variables. Word factors explained 60 percent of variance, character factors, 50 percent, and sentence factors, 12 percent when taken alone; 64 percent when all three were considered together.
Danish		
Lix Committee (Jakobsen, 1971)	Lix = M1 + Lo M1 = average length of meaning, i.e., sentence length Lo = percentage of long words, i.e., words with more than 6 letters In texts below 3,000 words whole text is lixed. For larger texts, there is a procedure for spot checks.	Based on Björnsson's Swedish research and the work of an official Danish Lix Committee convened in the 1960s and still active.

*Excludes research in which only cloze procedure was used.

Dutch

Douma, 1960	Uses Flesch formula, which estimates words and sentences 10 percent longer in Dutch than in English. Ease = 206.84 - 0.77sw - 0.93ws sw = syllables per 100 words ws = words per sentence	Generalized from five texts. Flesch's two coefficients are reduced by 11 percent.
Brouwer, 1963	Uses average length of words and average length of sentences as indices of difficulty. Places the two indices on the same footing. Ease = 195 - $^2/_3$ sw - 2ws	Criteria developed from study of 25 children's books.
van Hauwermeiren, 1972	L = 109.549 - 29.971 x_1 - 0.986 x_6 + 0.967 x_{10} L = readability level x_1 = average number of syllables per word x_6 = average number of nouns per 100 words x_{10} = average number of auxiliary verbs per 100 words	Validity = .65. Five other formulas exist, with validities from .60 to .67. Criterion – cloze.
Zondervan, van Steen, & Gunneweg, 1976	Grade 3: 6.44x_1 + 5.42 Grade 4: 6.58x_1 + 40.68 Grade 5: 5.76x_1 - 2.86x_2 + 45.67 Grade 6: 6.07x_1 - 3.20x_2 + 41.64 x_1 = percentage of different, difficult long words x_2 = percentage of auxiliary verbs Difficult long words are those with more than 3 syllables not on list of 35 most frequently used Dutch long words.	Criteria – cloze. Based on nonfiction texts for grades 3, 4, 5, 6.

Author and Date of Publication	Formula or Research	Observations
Staphorsius & Krom, 1985	Index for manual computation. $0.798 - 0.329\,\text{GWLG} - 0.004\,\text{GZLG} + 0.588\,\text{WW} + 0.472\,\text{SUB} + 0.654\,\text{PERS}$ GWLG = mean length of words in syllables GZLG = mean length of sentences in syllables WW = proportion of verbs SUB = proportion of substantives PERS = proportion of personal pronouns Index for machine computation. $1.284 - 0.15\,\text{GWL} - 0.010\,\text{GZW}$ GWL = mean length of words in letters GZW = mean length of sentences in words	Criteria – cloze. Recommended for nonfiction texts in grades 3, 4, 5, 6.
French		
Tharp, 1939	Proposes an Index of Difficulty in which the frequency index of a piece of reading material is divided by the density. Density is obtained by dividing the number of running words by the number of burden words.	Contrasts burden words with gift words, i.e., cognates and proper nouns. Stresses value of basic word lists for authors of second language texts.
Kandel & Moles, 1958	Adaptation of Flesch Reading Ease $\text{Ease} = 207 - 1.015\text{ws} - 0.736\text{sw}$ ws = number of words per sentence sw = number of syllables per 100 words	Because French words on the average are longer than English words, the coefficient for sw is divided by 1.15. Counting procedure not adapted to French.

De Landsheere, 1963	Flesch formula with coefficients unchanged but ways of counting specific to the French language.	Flesch formula computerized.
De Landsheere, 1966	Lexical base from Verlée word frequency list. Syntactic base a function of the way punctuation divides text.	Technique abandoned because of inconsistencies among authors in use of punctuation. Replaced by a more economical method using the Gougenheim word list.
Henry, 1973	Three sets of formulas 1. 8 Variable version requiring a knowledge of linguistics. 2. 5 Variable version ideal for computer. 3. 3 Variable version using number of words per sentence number of words absent from the Gougenheim word list first names used alone + exclamation marks + quotation marks Graphs available for use with third formula.	Criteria – 5 parallel cloze tests each from 60 books levels 1 through 12. Primary levels eliminated later. Three formulas in each set, one each for levels 5-6, 8-9, and 11-12. Interpolation possible for other levels. Not recommended for primary levels.
Richaudeau, 1973	Number of words retained from a clause = A + B + C A = Score from plotting intersection of average number of letters per word & average number of words on a graph. B = Addition of 2 points if sentence begins with an affirmative formula (This is why, etc.). C = Subtraction of 2 points when distance between subject and verb is more than 10 words.	Experimental formula based on the idea that the more readable a text, the more easily it is retained.

Author and Date of Publication	Formula or Research	Observations
German		
Fucks, 1955	Difficulty level = sentence length × word length. Based on a difficulty scale.	Judged inappropriate for German probably due to the frequency of long words in that language.
Walters, 1966	$Y = 801.12 - 40.77 (F_{87}) - 172.32 (F_{113})$ Y = difficulty index with range from approximately 100 (very difficult) to 550 (very easy) F_{87} = average number of verb segments per sentence F_{113} = density of modification in nominal units	Criterion – subjective judgment by author and 38 others of 15 300-word theological texts. Special purpose formula for use with theological literature in German. 33 formulas or their variations developed.
De Landsheere, 1970	German version of Flesch formula using same principles for counting as De Landsheere's French language version.	
Briest, 1974	Verb intensity = number of words divided by number of sentences.	Never used practically.
Schwartz, 1975	German Readability Graph similar to Fry Graph. Along horizontal axis, number of syllables per 100 words ranges from 125-189. Along vertical axis, number of sentences per 100 words range from 2.0-20.0. Predicts to grade 8+. Shows little difference between materials for grades 3 and 4.	Criteria – 100 word samples from West German readers of postwar era for grades 1-8; 15-21 samples for grades 2-8; and 11-15 samples for grade one. Average number of syllables per 100 German words greater than English by 25-37. Number of sentences close for corresponding levels.

Nestler, 1977	Conceptual level of words used to estimate difficulty level of texts. Three levels identified: 1. Generally known words. 2. Rare words. 3. Rare professional words.	Too difficult to use with complete passages. Only feasible with single sentences.
Dickes & Steiwer, 1977	Developed 8 variable, 6 variable, and 3 variable formulas. Three variable version is similar to Flesch instrument.	Multiple correlation with cloze scores on 60 German texts of .91, .89, and .87 respectively.
Bamberger & Vanecek, 1982, 1984	Developed a number of formulas, most requiring use of word list of 1,000 most common words in written language of a German speaking 10 year old. Exception is 4.WSF (fourth Viennese formula for nonfiction). Grade level = 0.2656sl + 0.2744ms - 1.6939 sl = sentence length ms = multisyllabic words Also devised subjective and other objective methods for evaluation of fiction and nonfiction. Subjective = checklist for evaluating 5 nonlanguage variables: content, organization, print, style, and motivation. Objective = separate profiles of language variables for fiction and nonfiction using additive techniques.	Criteria – 120 children's story books and 200 nonfiction juvenile books arranged into grade levels through use of pooled subjective assessment and by applying readability formulas used in development of profiles of language variables. Correlation with criterion of 4.WSF = .9724.

Author and Date of Publication	Formula or Research	Observations
Hebrew		
Nahshon, 1957	$GS = .236x_1 + .1338x_2 - 3.305$ GS = grade level at which an Israeli student can comprehend a passage without assistance x_1 = percentage of different hard words x_2 = average sentence length in words	Eight readability formulas developed for Hebrew prose. This is shortest with correlation of .868.
Hindi		
Bhagoliwal, 1961	Applied Johnson (1930), Flesch Reading Ease (1948), Farr-Jenkins-Paterson (1951), and Gunning (1952) formulas to 31 short stories in Hindi.	No Hindi word lists available, therefore limited to formulas involving syllable counts. Found Farr-Jenkins-Paterson best as it does not involve count of polysyllabic words, a problem in Hindi.
Korean		
Park, 1974	Multiple regression equation with five variables: easy words, different words, different hard words, simple sentences, and pronouns. Geared to materials for grades 2-9.	Criteria – graded language and social science books required for Korean schools by the Ministry of Education. Formula found more predictive for samples at lower grade levels.
Russian		
Rock, 1970	Readability graph based on the compilation of vocabulary items appearing in at least half of the Russian high school textbooks used in the U.S. and the percentage of unknown words that will result in the understanding or misunderstanding of authentic Soviet written text as demonstrated by research.	Preliminary study showed the acquisition of vocabulary is more difficult for American students in Russian than in German or Spanish. Result is slower development of proficiency in reading authentic materials in Russian than in the other two languages.

Spanish

Spaulding, 1951, 1956	$\text{Difficulty} = 1.609(\text{ASL}) + 331.8(\text{D}) + 22.0$ ASL = average sentence length D = density, based on the 1,500 word Buchanan word list Graph also available on which to plot variables. Rating is on a scale of from 20 (exceptionally easy) to 200 (exceptionally difficult).	Reliability for this formula is .87. More complex earlier one exists. This version widely used in Latin America.
Fernández Huerta, 1959	Adaptation of Flesch Reading Ease. $\text{Ease} = 206.84 - 0.60\text{P} - 1.02\text{F}$ P = number of syllables per 100 words F = number of sentences per 100 words	Tried Kandel & Moles Flesch adaptation for French first but found it had limited application.
Gutiérrez, 1972	$\text{Readability} = 95.2 - 9.7(\text{L}/\text{W}) - .35(\text{W}/\text{S})$ L = number of letters W = number of words S = number of sentences Validated at sixth grade level only.	Criteria – results of cloze tests administered to students in grade 6. Score yielded is in terms of average percentage of answers to a cloze test which students at this level would get on the passage being evaluated.
Patterson, 1972	Elaboration of Spaulding formula to help religious workers simplify written materials for readers with minimal reading skills.	
Thonis, 1976	Established grade levels for Spaulding's formula based on Patterson's descriptions of the reading ability needed to understand materials yielding various indices on the Spaulding scale.	Shortage of substantiating evidence makes grade level equivalency questionable.

Author and Date of Publication	Formula or Research	Observations
Garcia, 1977	Adaptation of Fry Readability Graph modifying horizontal and vertical axes to reflect differences in syllable and sentence length. FGAS (Fry Graph Adapted to Spanish).	Criteria – basal reading series in Spanish. Geared to English as a second language.
Gilliam, Peña, & Mountain, 1980	Adaptation of Fry Readability Graph retaining count for average number of sentences and subtracting 67 from average syllable count before plotting it on the graph.	Criteria – 13 textbooks and 9 juvenile books written in Spanish for use in grades 1-3; publishers' grade level on English version available and assumed to be on same readability level as Spanish text. 18 books had same readability in both languages. Suitability of graph for primary materials only evaluated.
Vari-Cartier, 1981	Adaptation of Fry Readability Graph increasing syllable count, altering sentence count, and changing readability designations to reflect the four general levels of second language study: beginning, intermediate, advanced intermediate, and advanced. FRASE (Fry Readability Adaptation for Spanish Evaluation).	Criteria – 127 samples from 66 American textbooks for Spanish language instruction. FRASE graph designations correlated with subjective teacher judgments, Spaulding formula ratings, cloze test scores, and informal multiple choice tests in a range of from .91 to .97.

López Rodríguez, 1981, 1982	Index of Difficulty = $95.4339 - 0.0756x_1 + 0.2012x_4 - 0.0669x_{16} - 0.0728x_{19} - 35.202x_{21} - 1.0601\ x_{22} + 0.7783x_{26}$ x_1 = commas x_4 = period and new paragraph x_{16} = words per sentence x_{19} = words of more than 3 syllables x_{21} = measure of redundancy x_{22} = Common Vocabulary of García Hoz x_{26} = expanded list of Spaulding	Criterion – cloze. Multiple correlation with criterion of .5618. Developed two more formulas for grades 7 and 8 separately.
Rodríguez Diéguez, 1983	Index of Difficulty = $59.929 - 0.098x_1 - 0.321x_2 + 4.428 \log (x_4) + 0.108x_{11} + 0.200x_{12} - 7.079 \log (x_{16}) - 25.816 \log (x_{21}) - 0.007 (x_{22})2 - 0.012x_{25} - 0.126x_{27} - 20.420x_{28} + 5.502x_{31}$ x_1 = commas x_2 = semicolons x_4 = period and new paragraph x_{11} = proper names x_{12} = numerals x_{16} = words per sentence x_{21} = measure of redundancy x_{22} = Common Vocabulary of García Hoz x_{25} = personal pronouns x_{27} = total of periods x_{28} = deviation of the distribution of letters per word x_{31} = mean of letters per word + 2.58 deviations	Criterion – cloze. Multiple correlation with criterion of .716. Added 8 variables to the 26 studied by López Rodríguez in search for his formula.

Author and Date of Publication	Formula or Research	Observations
Crawford, 1984	Readability graph based on the following regression equation Grade level = [number of sentences per 100 words × (-.205)] + (number of syllables per 100 words × .049) - 3.047	Criteria – 186 passages from Laidlaw series of Spanish readers, grades 1-6. Average sentence length and number of syllables per 100 words tabulated, their mean and SD calculated, and a multiple regression analysis of the data performed.
Swedish		
Björnsson, 1968a, 1983	Lix = average number of words per sentence + percentage of long words Long words = those with seven or more letters Twenty 100 word samples each for word length and 10 samples for sentence length recommended for SD of only 1.0. Rating on scale of from 20 (very easy) to 60 (very difficult); converted into grade levels for some languages by other researchers.	Criteria generally used for Swedish and other languages – approximately 100 varied passages whose difficulty has been rated by 20-30 persons. Validity of .95 on text evaluated when recommended number of samples used.
Platzack, 1974	Studied influence of punctuation marks, use of word order, relative pronouns, and sentence length on readability.	Research based on generative transformational grammar.
Vietnamese		
Nguyen & Henkin, 1985	RL = 2WL + .2SL - 6 WL = average word length SL = average sentence length Compound words counted as one word. Each tonal mark, word mark, and hyphen (in a compound word) counted as a letter. Readability table and scale available for easy computation. Readability given in terms of grade levels.	Criteria – 20 passages of approximately 300 words each from Vietnamese novels, magazines, and textbooks from grades 4 through college.

5

Edward B. Fry

Writeability: The Principles of Writing for Increased Comprehension

Readability formulas are concerned with judging the difficulty levels of writing. Frequently, textbooks, consumer contracts, and a wide variety of reading matter are found too difficult or unreadable for the intended audience. *Writeability* is concerned with writing, rewriting, or editing to get those materials to the desired readability level.

Writeability helps writers and editors produce materials that can be comprehended more easily without cheating. *Cheating* is defined as trying to beat the formulas by artificially chopping sentences in half and selecting any short word to replace a long word. You can cheat on an IQ test to get a higher score, but cheating will not change your intelligence. Likewise you can cheat or artificially doctor writing to get a lower readability formula score, but you might not have changed the true readability much and you may have made it worse.

True readability is the goal of most authors. They want to communicate ideas to the reader. Increased readability can be demonstrated by higher scores on comprehension questions, ability to fill in more blanks in a cloze passage, fewer oral reading errors, a tendency to spend more time reading, more mature eye movements, and subjective judgment of the reader.

Despite criticisms of readability formulas, their use has never been more popular. The formulas have had a profound influence on the textbook publishing industry in the past ten years. Most editors and sales personnel can tell you the readability score of their materials. The influence on technical materials has been equally great, and readability formulas have influenced the writing of such diverse products as army maintenance manuals, insurance policies, and income tax booklets.

New York, New Jersey, Hawaii, Connecticut, and Minnesota have passed Plain Language Laws. California and Michigan are actively considering joining the group. In Oklahoma the State Department of Education checks ballot propositions for readability.

Plain Language Laws differ from state to state but their basic goal is to simplify all types of consumer contracts such as rent agreements, money lending forms, and the fine print of insurance policies. The New Jersey law states that "a consumer contract shall be simple, clear, understandable, and easily readable."

In 1978, President Carter sent Executive Order 12044 on plain language to all government departments; many states have similar movements.

Basing reading tests on a readability formula is important. The Degrees of Reading Power now taken by every student in New York state and used in Boston and parts of Connecticut is not a norm based (standardized) test. Rather, it yields a readability score. In other words, it tells the teacher or administrator what books the students can read. It does not compare one student with another, though it could be used in that manner.

The basic idea behind readability has always been to help writers, editors, teachers, and librarians to match the difficulty of written material with the reading ability of the student. A good match improves communication and learning.

This article will discuss how readability formula scores can be lowered without cheating.

Vocabulary

Since a major input of most readability formulas is vocabulary difficulty, one way to lower readability scores is to use simpler vocabulary.

Writers and editors can use word frequency lists as guides. The American Heritage list (Carroll, Davies, & Richman, 1971) that ranks 87,000 words found in frequency counts of 5 million running words has largely replaced the older but still valuable Thorndike-Lorge lists. Most writers will find these lists too expensive and cumbersome. A shorter and more usable list is Sakiey and Fry's (1979) list of the 3,000 most used words in English, both in rank and in alphabetical order. The commonness of the variant forms (adding *s, ed, ness*) is given for each word. Even shorter word lists like the Dolch 220 Basic Sight Words or the first 300 Instant Words can be useful to writers of primary material.

The first rule for writers is to use more common (high frequency) words. For example, don't "prolong a process"—"keep it short." Take care not to substitute words of Latin or Greek origin for common words. For example, *proceed* often means *go,* and *secure* often means *safe.* Words beginning with *pre, dis,* or *multi* often can be substituted for easier words.

Much of what is commonly called jargon or gobbledygook is simply someone using large words to sound pretentious or self-imposing. Readability formula makers are aware that there are times when the longer word is necessary and should be used. Longer words can add precision, clarity, and grace to an author's writing. But use too many and you will get a truly higher readability score. Changing the frequency of 15 percent of the words in a sixth grade basal reader story, for example, significantly increased reading comprehension performance (Marks, Doctorow, & Wittrock, 1974).

A group of science textbook editors wanted to use the word *temperature* in a primary book and *osmosis* in an upper level book. They argued that it is awkward not to use the correct word, and, furthermore, students should learn to read those words. I agreed, and a modification of my formula appeared in *Publishers Weekly* (1979), and is restated here. Any term presumed to be new and difficult for the readers should:

1. Be defined or used in context the first time it appears in such a way that its meaning is apparent and
2. Be followed for the next three times it appears by the pho-

netic spelling or syllabification of the word as found in any commonly used dictionary. For example (syl lab i fy) or (sĭ lăb′ə fī).

3. The new term now can be counted as a two syllable word regardless of its length.
4. These new terms should appear in the teacher's guide; as a new word list at the beginning or end of a chapter in the students' book; and, possibly, in a glossary.

In brief, the author is teaching the term to the reader in a helpful, meaningful way. I'm in favor of vocabulary improvement; I'm just opposed to bad communication that occurs when the author uses words the reader does not know.

Another important type of word list is based on words known by students at given grade levels. The most impressive of these is the *Living Word Vocabulary* (Dale & O'Rourke, 1976), which lists meanings of words known at different grade levels. For example, the word *run* has many meanings. At grade four, students know it as a baseball word; at grade six, the way a political candidate uses it; at grade eight, as the way to manage a business, and at grade twelve, as a sudden demand. No present readability formula takes this degree of complexity into account, though some formulas (such as the Dale-Chall) do count as unfamiliar or difficult those words not known by a majority of fourth graders. Meaning lists are valuable as a writer's resource.

Sentences

The other major input to most readability formulas is a measure of syntactical complexity. To put it briefly, readability formulas measure average sentence length. They could measure grammatical constructions such as prepositional phrases or subjunctive clauses, but most of these measures correlate highly with average sentence length. Hence, the obvious instruction to authors is *keep your sentences short*—on the average.

Nothing is more boring reading than a long series of short choppy sentences. On the other hand, nothing makes writing less understandable than very long sentences. Variety is necessary to express yourself properly and to interest your reader. Short sentences hit hard. Longer sentences subtly suggest that the idea simply cannot be expressed without some reservations and qualifications.

Historically, sentences are getting shorter. Flesch (1974) reports that Elizabethan sentences averaged 45 words, while Victorian sentences averaged 29 words. He found that sentences in magazines such as *Time* and *Reader's Digest* averaged 18 words in 1949 and 15 to 17 words in 1974. Flesch estimates that there was a 10 percent shortening in sentence length in twenty-five years, a rather hefty shrinkage. The American Press Institute conducted a Reading Comprehension Survey of 410 daily newspapers and found a strong correlation between words per sentence and reader comprehension. In most cases, the shorter the sentence, the more easily it was understood.

Just shortening sentences is not the total answer. Critics of readability formulas have pointed out that sometimes longer sentences communicate better. For example, you might break a long sentence like *Farmer Brown didn't go to town because the roads were icy* into two simple sentences as *Farmer Brown didn't go to town. The roads were icy.* If you then asked students, "Why didn't Farmer Brown go to town?" you might find that more of them got the answer correct after reading the longer sentence. This is why readability formulas say that "on the average" sentences should be shorter for better communication. They do not say that every sentence should be short.

Some longer sentences are said to add cohesion to writing (Kintsch & van Dijk, 1978; Kintsch & Vipond, 1979). An example of this is the "because" sentence used earlier about Farmer Brown. Other examples of longer sentences that could add to easier reading are if/then and either/or sentences. But in general, the basic readability principle is that shorter or less embedded sentences are easier to read.

Remember, readability formulas are not meant to be writer's guides. They are meant to judge the difficulty of a prose passage *after* the material has been written. This article is a writer's guide.

There are two more important kinds of sentence complexity that readability formulas cannot pick up. The first is the Kernel Distance Theory. It states that splitting the subject and verb-object (the kernel of the sentence) with distance (extra words) causes poor readability. An example of a split kernel would be: *Children,* if they don't wish to get colds, *should wear mittens."* An unsplit kernel would read *"Children should wear mittens* if they don't wish to get colds." This theory also states that distance in front of the kernel is worse than distance at the end of the kernel. For example, "If they don't wish to get colds, *children should wear mittens."* Note that all three sentences have the same vocabulary and the same sentence length and would get the same readability score, but research has shown that split kernel sentences yield worse communication. Other writers would call kernel splitting "embedding."

A second factor, often mentioned in rhetoric books, is that active sentences communicate better than passive sentences. Don't say "The test was taken by the students." Say, "The students took the test." Sometimes, as in this case, the active sentence is shorter. Klare (1980) suggests that writers should use active verbs rather than nominalizations (verbs made into nouns). For example, the verb *to sign* can be nominalized as *signature.* This leads to indirect writing like "Your signature must be affixed to the form." It is better to say "You must sign the form."

Most of the time, punctuation is helpful to the reader. A notable exception is the overuse of commas. Sprinkling too many commas in a sentence means that the flow is choppy and may also indicate that you have a heavily embedded sentence or one that is too long. Too many commas warn the writer or editor that something may be amiss, and so does the use of semicolons or colons.

Paragraphs

On the average, paragraphs should be short. Paragraphs are intended to guide the reader into seeing units of thought, gestalts, or schemata. Whatever they are called, they should have some kind of psychological units of cohesion. A very long paragraph often contains too many different ideas; short paragraphs have punch.

The traditional English textbook admonition about writing a well structured paragraph with main idea, supporting details, and conclusions is partly a myth. It might help to train neophyte writers, but it simply isn't followed by many professional writers. Take a look at Hemingway or any newspaper and you will see plenty of short, even one sentence, paragraphs. There are times when you will want to use a large, well structured paragraph, but don't be too bound by some formal ideal of paragraph structure.

Another factor related to paragraphing is the use of lists. They are particularly useful in technical writing or in directions. Lining up a list of terms or objects is often better than stringing them all together in a sentence, separated by commas. This is another instance of too many commas being a danger signal.

Organization

Selecting the proper organization for an article or a chapter is part of the art of being a writer. Some subject matters lend themselves more to one form of organization than to another. For example, history often relies heavily on a chronological organization, but effective theme or problem centered histories have been written.

One type of organization effective in expository writing is the Statement-Example-Restatement (SER) sequence. SER includes repetition, giving concrete examples, and restating the principle in another way.

Subheads contribute to understanding an article and to informing the reader what organizational pattern the writer is using. Good subheads can act a little like Ausubel's advance organizers, Rothkopf's interspersed questions, or Kintsch's discourse pointers. Subheads also help the overview and review processes recommended in the SQ3R and other study skill techniques. They are used by skilled readers to improve comprehension or retention of the material.

Clearly written materials use many signal words to indicate the author's organization to the reader. Signal words can indicate (1) sequence and rank order such as *first, second, next, last, in conclusion;* (2) that a reverse idea is coming—*however, but, on the other hand;* and (3) that the author is not absolutely certain—*maybe, if, allegedly, might.* Other words signal that an example is coming up or that ideas might be paralleled or a choice given.

Cohesion

Cohesion refers to how well a paragraph or a passage "hangs together." Disjointed sentences or paragraphs indicate lack of cohesion. Cohesion was illustrated in the "because" sentence earlier to show that ideas or concepts sometimes can be communicated better in longer sentences. The "because" tied two ideas together to make them more cohesive.

On a connected discourse level (units of prose longer than one sentence) the SER is another type of cohesion where different parts of the story are repeated and interrelated. Signals words and summaries help tell the reader how the article is cohesive.

A number of current researchers such as Halliday and Hasan (1976), Kintsch and Vipond (1979), and Meyer (1977) have been concerned with cohesiveness, analyzing propositions (single thoughts, ideas, or concepts), and showing how they are related. They use terms like *links*, *ties*, and *networks*. In general, they would argue that cohesion is aided by more links between propositions. Hence, moderate use of referents (though not too distant) is good.

Traditionally, paragraphs should be cohesive by being about a single thought. In modern terminology, paragraphs should be cohesive by having the propositions well linked which means, "Do not jump from idea to idea too quickly." Many different ideas in a short passage make readability difficult. Cohesion extends beyond the paragraph to entire books.

Personal Words

The American Psychological Association condones and, at times, encourages the use of personal pronouns in scholarly writing. For example, many reports conclude with "It was found that...." Who or what is *it*? Chances are that *it* is none other than the author who prefers a literary castration rather than the use of a personal pronoun, thinking it is more scholarly. Not using personal pronouns results in poorer communication and borders on academic dishonesty. What really happened was "I found that...," so why not say so. Take personal credit or blame, and be a real person to your reader.

For those who need more convincing, here is a direct quotation from the APA *Publication Manual:*

> An experienced writer can use the first person and the active voice without dominating the communication and without sacrificing the objectivity of the research.

The APA Manual also cites the *American National Standard for the Preparation of Scientific Papers:*

> When a verb concerns an author's belief or conjecture, use of the impersonal passive ("it is thought" or "it is suggested") is highly inappropriate. When a verb concerns action by the author, he should use the first person...."

However, do not use too many personal pronouns because they draw attention away from the subject and toward the author.

Flesch (1974) developed a formula for measuring interest by using personal words and personal sentences. It is not as well known as his readability formula, and the concept is not as widely accepted.

Personal sentences are those sentences aimed directly at readers. For example, "You should always...." Another type is the sentence used in dialogue with direct personal reference. For example, "Sally said...."

Personal words and sentences apply not only to adult writing; they also are important in children's writing. When sources as diverse as Rudolf Flesch and the *APA Manual* encourage you to use personal pronouns and personal sentences, you should consider doing so.

Imageability

Imageability refers to the ease with which the reader can visualize the word, phrase, or whole passage. Some writers refer to this as concreteness. A number of psychological studies, such as those by Paivio (1969), have found that highly imageable words are easier to learn and remember. For example, *dog*, *bulldozer*, and *mother* are high imageable words and *is*, *philosophy*, and *of* are low imageable words. In the medium range you might find *blue*, *running*, and *under*.

Phrases, sentences, and whole passages can be ranked for imageability. Good writers are adept at finding vivid examples. Abstract statistics are exemplified by a description of a typical person or product that illustrates the mean or modal findings. Writers of children's science books are often ingenious in illustrating basic principles of physics with familiar situations. Pastors use stories from real life to illustrate the ten commandments, and educators attempt to make teaching principles more understandable by citing actual classroom incidents.

Metaphors are an attempt to increase imageability. They often attempt to give a concrete corollary to a less visual concept. For example, "Her personality began to unfold like a rose. It was a hard tight bud the first day on the island, but each morning after the sun arose it opened a bit more until the full flower of womanhood was revealed."

You can improve imageability by adding appropriate pictures, diagrams, maps, and graphs to your manuscript.

Referents

Improper use of referents makes some writing hard to follow. *Referents* (sometimes called anaphora) can be pronouns (such as *it, they,* and *their)* or phrases. For example, "the old man" can refer to Captain Ahab. Referents are words that must refer to something, and that something must be clearly understood, usually by having been used in the preceding sentence.

Writers use referents because they save time; it isn't necessary to continually repeat the full noun. For example, it is a little awkward to keep repeating *The Lord High Executioner* when in the next sentence you can say *he.*

The misuse of anaphora causes trouble; for instance, when the referent can refer to more than one thing. *They* might refer to the good guys or the bad guys and it makes a lot of difference. A greatly delayed referent can impose a burden on reader memory and a momentary loss of comprehension or story flow. The way out of these difficulties is simple – repeat the noun.

Older lawyers were famous for jargon and greatly delayed referents. Early in the document they might use the phrase "hereinafter referred to as the party of the first part." Pages later, the reader must remember who was the party of the first part. Modern word processors have a simple solution. Simply program the word processor to insert "Jones" every time you come to "party of the first part."

Motivation and Subject Matter

One thing readability formulas will never be able to judge is an individual student's motivation to read a particular bit of writing. Some time ago, I proposed The Readability Principle. Briefly summarized it says: High motivation overcomes writing that is hard to comprehend. What secondary teacher has not witnessed a teenage male student with sixth grade reading ability (according to nationally standardized tests) master a driver's license manual written at tenth grade readability? This same teenager has difficulty comprehending his social studies book coupled with reluctance to read it for very long.

Writers who want to be read should find interesting topics. If they must write on a difficult topic, they should seek interesting examples and applications.

This goes along with the general injunction to know your audience. Write directly to someone. Select the proper level of sophistication, then try to write a little below that level. Best selling novels are written at an eighth grade level according to readability formulas, but they are read mostly by high school graduates. The reason is not that novel buyers are semiliterate or that they can read only at the eighth grade level, but rather as readability formula makers have said all along "lower readability scores mean there is an inclination on the part of the reader to continue reading the material." If business executives want their memos read or manufacturers want their instruction manuals read they need to keep the readability scores low.

Try to be aware of your reader's background knowledge. What does the reader bring to the text? You can assume some intended audiences are familiar with the concept and a brief mention is all that is necessary. For other audiences, more explanation is necessary. This idea of background knowledge is related to vocabulary difficulty. Does the audience know the words you are using? Even beyond vocabulary, the writer should know where the reader is coming from.

Try Outs

There is one sure way to find out if you have achieved true readability. Try the writing on a sample of the intended audience.

You can try to follow all the principles discussed in this article, you can have the article edited or reviewed by peers or superiors, but nothing proves readability like a try out. If you are writing for children at a certain grade level, get an average student, a slightly below average student, and a slightly above average student, and have them read it. Do not go to extremes and use very bright or very poor students.

The same is true in writing for adults. If you are writing contracts, ballot propositions, or even newspaper articles try out samples of your writing on a few members of your intended audience.

Check the comprehension of your try out sample by discussion and formal or informal questions. If the writing is a set of directions, see if your sample audience can follow the directions.

If you haven't communicated effectively, the material needs to be rewritten or edited. The ideas from this article are summarized on the Writeability Checklist. It can help you, particularly in rewriting or editing.

"Easy reading is hard writing" is a principle writers have known for a long time. But good writing to an appropriately easy level can be improved with practice, and try outs on a real audience are an excellent source of feedback.

Legal Status

In addition to the Plain Language Laws enacted by some states and to the Presidential Executive Order, a number of court cases have dealt with readability. For example, a man was blinded by a drain cleaner when he did not follow directions written on the can.

A large scale class action suit brought in Federal Court against Medicare (Weinstein, 1984) used a readability formula as a pivotal instrument. There are some 50,000 Medicare claims filed every day and since Medicare pays an average of only 62 percent of the claimed amount, it is not surprising that some 1,500 appeals are filed. According to the Fry Readability formula, the notice sent to these review seekers was found to have a readability ranging from twelfth to sixteenth grade. Couple this with the fact that 48 percent of the citizens of New York who are age sixty-five or older have an eighth grade or less education. Chief Judge Jack B. Weinstein found that the notices sent out by Medicare were "incomprehensible" and that these "inadequate notices can be remedied. Defendant [Medicare] is directed to take prompt action."

It Can Be Done

There is ample evidence that proper writing or rewriting can keep the readability of most materials lower than is commonly supposed. Recently, a student rewrote part of the New Jersey Drivers License Manual as part of a master's thesis. When the readability score was lowered two grade levels from eighth to sixth grade, cloze scores jumped significantly. A surprising finding was that even though the rewritten passages were longer, the students completed the cloze tests in less time, indicating that if a passage is easier to read it can be read more rapidly (Hunt, 1982).

The Document Design Center (1982) reports a field test of an FCC regulation written in original bureaucratic style and a version rewritten to be more readable. Readers' responses were more accurate and faster when the more readable version was used.

Another of my students applied a readability formula to a front page story in the undergraduate student newspaper and to a front page story in the *New York Times*. He was amazed to find that the undergraduate written story was at the seventeenth grade level and the *New York Times* writer wrote at the eleventh grade level.

Too many people write like that undergraduate, and it simply isn't necessary.

As an interesting summary of many of the points made in this chapter, the reader might like to see some advice to writers given by C.A. McKnight, an editor of the *Charlotte Observer:*

1. Use short, simple words. If your writing runs more than 165 syllables per 100 words, you are writing only for college graduates.
2. Use more one syllable words. Make them your workhorse words. Make them carry the biggest load. Of 275 words in Lincoln's Gettysburg Address, 196 are only one syllable.
3. Use familiar words. The Bible uses a vocabulary of only 6,000 words.
4. Use personal words. Your stories will come to life when you sprinkle in a generous supply of words such as *you, girl, mother, doctor, teacher, Joe, Susie, baby.*
5. Use concrete words—words that make the reader see, hear, feel, smell, or taste.
6. Make every word work. Use fewer words. Use them with greater force. Go through one day's writing after it's printed. Cross out every unnecessary word, confusing phrase, garbled sentence, involved paragraph. Then continue to do that every day, in advance of printing.
7. Avoid technical words. Nontechnical words are clearer, and will build a broader base of readership for you.
8. Create figures of speech. Build them into everyday writing. Feed new ones in as old ones wear out.
9. Use short sentences. They are the lifeblood of simple, easy to read writing. If a sentence runs upward of 30 words, break it up. Even a one word sentence can be forceful, emphatic, arresting.
10. Make sentences active. Put a taboo on passives. Active verbs give action to writing, passives bog it down.
11. Use short, simple paragraphs. Most should introduce or contain only one idea. Most should also have only one source or viewpoint.
12. Write to one person. Write every story or feature as if you were talking to one man, to one woman, to one child. Picture this person sitting right in front of you as you talk. Talk to that person in familiar language, words used every day.

13. Work with one basic idea. Cover many points, but build them on the framework of one idea. That can make even a complex subject easy to read about.
14. Try to write affirmatively. Keep your viewpoint constructive. There is always a *yes* viewpoint in every *no* situation.

(American Press Institute, 1985)

In case you prefer to learn from the negative instead of the positive, here is a delightful quote from Law Professor Robert Benson (1984-1985):

> There exist scores of empirical studies showing that most of the linguistic features found in legalese cause comprehension difficulties. Legalese is characterized by passive verbs, impersonality, nominalizations, long sentences, idea stuffed sentences, difficult words, double negatives, illogical order, poor headings, and poor typeface and graphic layout. Each of these features alone is known to work against clear understanding.

You might note that of the eleven negative characteristics of poor writing, readability formulas take into account only two. This is why readability formulas are not writer's guides.

Instead of a summary, you are invited to look over Appendix 1, a Writeability Checklist that mentions most of the ideas contained in this article. The important point is that the checklist contains many more factors than the two simple inputs of a readability formula, sentence length and word length, as indicated by number of syllables or some other measure.

For those who do not have a readability formula, Appendix 2 shows the handy Graph for Estimating Readability. It will give you a fairly reliable estimate of the difficulty of any piece of prose, not with deadly accuracy, but with accuracy comparable with most other human psychological measures.

Any readability formula is meant to be used after a piece of prose is written. Probably the best advice for writers is to write creatively (some say there is no such thing as uncreative writing) aiming it at your intended audience. Then use a readability formula. If it comes out on the proper level, quit.

If your writing is not on the intended level, the Writeability Checklist will help you simplify it. Few people have trouble writing material which is difficult to read – sentence combining is a skill taught in elementary school. It takes art and talent to write in a simple, clear manner. Perhaps the Writeability Checklist will help you toward that laudable goal. At very least, the Writeability Checklist will mollify some of the critics of readability formulas who fear that formula makers want all writing to be just short choppy sentences and short choppy words.

References

American National Standard for the preparation of scientific papers for written or oral presentation. (ANSI Z39.16-1972). New York: American National Standards Institute, 1972.

American Press Institute. *Effective writing and editing.* Reston, VA: American Press Institute, 1985.

Benson, R.W. The end of legalese: The game is over. *New York University Review of Law and Social Change*, 1984-1985, *13* (3).

Carroll, J.B., Davies, P., and Richman, B. *The American heritage word frequency book.* Boston: Houghton Mifflin, 1971.

Dale, E., and O'Rourke, J. *The living word vocabulary.* Elgin, IL: Dome, 1976.

Document Design Center. *Simply stated*, Number 29. Washington, DC: Document Design Center, American Institute for Research, 1982.

Flesch, R.F. *The art of readable writing.* New York: Harper & Row, 1974.

Fry, E.B. Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading*, 1977, *21*, 242-252.

Fry, E.B. A reappraisal of formula writing. *Publishers Weekly*, October 29, 1974.

Halliday, M., and Hassan, R. *Cohesion in English.* London: Longman, 1976.

Hunt, S.M. *The communicative effects of a rewritten driver's manual.* Unpublished master's thesis, Rutgers University, 1982.

Kintsch, W., and van Dijk, T.A. Toward a model of text comprehension and production. *Psychology Review*, 1978, *85* (5), 363-394.

Kintsch, W., and Vipond, D. Reading comprehension and readability in educational practice and psychological theory. In L.G. Nilsson (Ed.), *Perspective on memory research.* Hillsdale, NJ: Erlbaum, 1979.

Klare, G.R. *A manual for readable writing.* Glen Burnie, MD: REM Company, 1980.

Marks, C.B., Doctorow, M.J., and Wittrock, M.C. Word frequency and reading comprehension. *Journal of Educational Research*, 1974, *67* (6).

Meyer, B. The structure of prose: Effects of hearing, memory, and implications for educational practice. In R.C. Anderson, R.J. Spiro, and W.E. Montague (Eds.), *Schooling and the acquisition of knowledge.* Hillsdale, NJ: Erlbaum, 1977.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, *76*, 241-263.

Publication manual of the American Psychological Association. Washington, DC: American Psychological Association, 1974.

Sakiey, E., and Fry, E. *3,000 instant words.* Providence, RI: Jamestown Publishers, 1979.

Weinstein, J.B. "Joseph David et al., Plaintiffs against Margaret Heckler, Secretary, Department of Health and Human Services, Defendants." Civil Action No. 79 C 2813, United States District Court, Eastern District of New York, Brooklyn, New York, July 11, 1984.

Appendix 1
Writeability Checklist

Vocabulary

______ Avoid large or infrequent words
______ For high frequency words use the Carroll list or *3,000 Instant Words*
______ For meaning list use *Living Word Vocabulary*
______ Avoid words with Latin and Greek prefixes/roots
______ Avoid jargon
______ OK to use technical words but see rules for introducing new terms

Sentences

______ Keep sentences short; on the average for general adults keep average sentence below 15 words
______ Avoid splitting sentence kernel (embedding)
______ Keep verb active (avoid nominalizations)
______ Watch out for too many commas (may indicate need for two sentences)
______ Semicolons and colons may indicate need for new sentence
______ Cohesion sometimes aided by longer sentences

Paragraphs

______ Keep paragraph short, on the average
______ One sentence paragraphs permissible at times
______ Indent and line up lists

Organization

______ Suit organization plan to topic and your purpose
______ Consider SER (Statement-Example-Restatement)
______ Use subheads
______ Use signal words
______ Use summaries

Cohesion

______ Increase links between sentences and paragraphs
______ Avoid too many different ideas in a short passage

Personal Words

______ Use personal pronouns, but not too many
______ Use personal sentences
______ A direct statement to reader or dialogue

Imageability

______ Use more high imageable words (concrete)
______ Avoid low imagery words; edit out many
______ Use vivid examples
______ Use metaphors
______ Use graphs whenever appropriate

Appendix 1
Writeability Checklist (continued)

Referents

______ Avoid too many referents
______ Replace some referents with the original noun or verb
______ Avoid distance between noun and referent
______ Don't use referents that could refer to two or more nouns or verbs

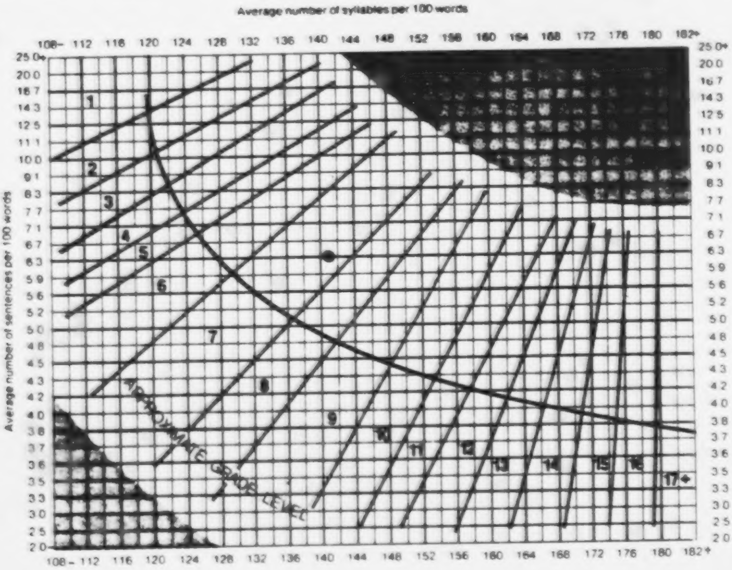
Motivation

______ Select interesting topics
______ Select interesting examples
______ Write at a level a little below your audience
______ Consider readers' background knowledge

Try Outs

______ Try out writing on a sample audience
______ Check comprehension by sample audience
______ Revise if necessary

Appendix 2
Fry Graph for Estimating Readability – Extended



DIRECTIONS: Randomly select 3 one hundred word passages from a book or an article. Plot average number of syllables and average number of sentences per 100 words on graph to determine the grade level of the material. Choose more passages per book if great variability is observed and conclude that the book has uneven readability. Few books will fall in gray area but when they do grade level scores are invalid.

Count proper nouns, numerals and initializations as words. Count a syllable for each symbol. For example, "1945" is 1 word and 4 syllables and "IRA" is 1 word and 3 syllables.

EXAMPLE:	SYLLABLES	SENTENCES
1st Hundred Words	124	6.6
2nd Hundred Words	141	5.5
3rd Hundred Words	158	6.8
AVERAGE	141	6.3

READABILITY 7th GRADE (see dot plotted on graph)

Reproduction Permitted. No Copyright.

Readability: Its Future

6

Marilyn R. Binkley

New Ways of Assessing Text Difficulty

As interest in textlinguistics increased during the late 1970s and early 1980s, researchers developed new descriptions of text structure. At the same time, researchers became increasingly disappointed with the ability of classic readability formulas to describe adequately the features of text that influence comprehension or to guide the production of improved text. Together, these trends are influencing reading comprehension research.

The intent of the new text analysis systems is significantly different from that of classic readability formulas. The new systems attempt to predict ease of reading, to test hypotheses about thought processes, and to guide production. Therefore, instead of focusing on factors of text correlated primarily with reading difficulty, the new approaches try to identify text factors that influence learning and memory.

Many of the new approaches have been developed by researchers in fields other than reading, particularly rhetoric (D'Angelo, 1975; Flower & Hayes, 1977), linguistics (Fillmore, 1968; Grimes, 1975; Halliday & Hasan, 1976; van Dijk, 1977), psychology (Bartlett, 1932; Dawes, 1966; Frederiksen, 1972; Mandler & Johnson, 1977; Rumelhart, 1975), and artificial intelligence (Charniak, 1972; Schank, 1977; Simmons, 1978). However, they have been useful to reading researchers because they make possible new ways of thinking about reading. They also have potential for measuring text difficulty.

The opinions and suggestions expressed in this paper are those of the author and do not necessarily reflect the positions or policies of the U.S. Department of Education.

This chapter provides a review of the most significant new text analysis systems, describing each type, discussing its usefulness to reading research, and exploring its applicability as a readability estimate. The chapter concludes with a proposed methodology that extends cohesion analysis into an ease of reading measure.

Defining Text and Language Structures

There are two key attributes in all text analysis systems: the unit of discourse and the kind of relationship.

Researchers have proposed a number of divisions of discourse. For example, van Dijk (1979) differentiates between macro and microstructures. Armbruster (1984) looks at global or local coherence. Meyer (1981) identifies three text levels: sentence, paragraph, and top level. These differentiations focus on how much of a text is being considered, how pieces of text relate, and which types of rhetorical structures are used in developing a text.

Similarly, at least three distinct sets of relationships operate within text. Grimes (1975) identifies the structures as content or semantic, cohesive, and staging. According to Grimes, these relationships interact in discourse, causing coherent text to form in such a way that the theme is selected from already introduced information and then related cognitively and thematically to the rest. However, each set of relationships may be studied separately and examined at either a micro or macrolevel.

Viewed in this broader context, we can categorize text analysis systems along the two dimensions noted—the unit of discourse and the kind of relationships examined. Figure 1 is a simple model of such a categorization.

The Figure illustrates that, in theory, text analysis systems can be devised to include any combination of study units and types of relationships. For example, Frederiksen's semantic networks system (described later) focuses primarily on the content structure at a microlevel. Not all such combinations have been fully explored. To date, most text oriented research has focused on content structure, which may be described at both the macro and microlevels. In con-

Figure 1
Categories of Text Analysis Systems

		Level of Discourse	
		micro	macro
Types of Relationships	semantic	networks propositional networks	story grammar rhetorical structures
	cohesive	cohesion analysis	
	staging		

trast, research based on the cohesive structure of text has been confined to the microlevel, due perhaps to a limited understanding of these systems. As work by Hasan (1980) attests, cohesion chains operate at both micro and macrolevels of text.

The Content Structure

Content structure refers to the cognitive structure of text – the semantic, or meaning, aspects. It most clearly represents the information and complexities of text.

Attempts to represent the content of existing passages include set relations (Dawes, 1966), linear relationship structures (Frase, 1973), propositions (Kintsch, 1974), and networks (Frederiksen, 1975a). These operate primarily on the microlevel. Another approach, constituency grammars, attempts to identify functional units between the proposition (microlevel) and the passage (macrolevel). Examples include story grammars and rhetorical predicates (Meyer, 1975; Rumelhart, 1975).

Frederiksen's (1975b) semantic networks and Kintsch's (1974) propositional analysis are at the foreground of content structure analysis. Both are based on Fillmore's (1968) notion of case, which assumes that deep structure or meaning does not vary with the surface structure of sentences. For example, consider two sentences:

John opened the door with the key.

The door was opened by John with the key.

Although the surface and syntactical structures differ, each of the noun phrases retains the same case or role relationship. By examining how various cases relate to the verb phrase and plotting this relationship, the meaning of a passage can be separated from its surface form.

This example from Kintsch's propositional analysis demonstrates how these systems work. Consider the simple sentence:

Mary is baking a cake.

In propositional analysis, this would be recorded as

(Bake A:Mary, O:cake)

where *A* is the agent and *O* is the object. Any surface structure that has Mary baking the cake can be reduced to this relationship. Therefore, when comparing various versions of the same story, the analyst can determine if the same propositions were represented.

Meyer and Rice (1984) have created an analysis system that combines the micropropositional elements of the Kintsch and Frederiksen approaches with a macropropositional element. They identify five types of rhetorical relations in text: causal, problem and solution, comparison, collection, and description. These rhetorical structures facilitate the segmentation of text into a hierarchical form that can be used to identify the top level structure of expository passages.

Researchers have used these three approaches to determine the effects of variation of structure on students' understanding, learning, and retention of content. The approaches work particularly well for representing text and comparing protocols to that text. As such, they are useful for determining learning in relation to a particular text.

Research using these systems of text analysis has led to findings that could have an impact on text production.

- Ideas located at the top levels of a structural analysis of prose are recalled and retained better than ideas located at the lower levels (Bartlett, 1978; Britton et al., 1979; Duchastel, 1979; Haring & Fry, 1979; Meyer, 1971, 1975, 1977; Swanson, 1979).

- Different items of information located high in the structure are more likely to be integrated in memory than items located low in the structure (Walker & Meyer, 1980).
- The type and structure of relationships among ideas in prose dramatically influence recall when they occur at the top levels of the structure; however, when the same relationships occur low in the structure they have little effect on recall (Meyer, 1975).
- Different types of relationships at the top levels of the structure differentially affect memory (Meyer & Freedle, 1984).
- Students who are able to identify and use these top level structures remember more from their reading than those who do not (Meyer, 1979; Meyer et al., 1980).
- Training in how to recognize and use these top level structures improves recall for text materials (Bartlett, 1978).
- Overgeneralizations, pseudodiscriminations, and text generated inferences occur at the time of comprehension, while elaborations occur during recall (Frederiksen, 1975).
- Explicit statements of logical relationships facilitate comprehension in poorer readers (Marshall & Glock, 1978).

However, these systems of analysis do not easily lend themselves to determinations of relative reading levels.

Story grammars, on the other hand, have led to experimental readability measures. Story grammar is based on the premise that a reader understands the organization and elements of a story, independent of the specific content. The story grammar represents the important elements in a story and specifies the allowable ways elements may be arranged (Black & Wilensky, 1979).

Many story grammars have been proposed. Almost all describe stories as consisting of a setting and a series of one or more episodes. Each episode tends to have an internal structure made up of a problem/solution or a goal/action/outcome. In addition to rewriting the story as a listing of rules, a reader may redraw a story grammar as a tree diagram that illustrates the hierarchical relationship among the constituent parts (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979). As such, the story grammar tends to have a top down or macroproposition orientation similar to Meyer's rhetorical predicates.

Meyer and Rice (1984) point out that, in reading research, a given story is analyzed so the components of the passage are identified according to their role in the story. Stories then can be compared based on their structure. This has led to some interesting findings.

- Recall is easier for a second story with the same structure as an earlier story (Thorndyke, 1977).
- Comprehensibility ratings of stories can be predicted (Bower, 1976; Rice, 1978).
- The sorts of summaries subjects will make of target stories can be described (Kintsch, 1977; Kintsch & van Dijk, 1975; Rumelhart, 1977).
- The items that will be remembered from a story can be predicted (Mandler & Johnson, 1977; Thorndyke, 1977).

In contrast to researchers aiming at understanding the reading process and story grammar acquisition, others have attempted to convert grammars into "a quantitative means of predicting the readability of a story." For example, Templeton and Mowery (1985) analyzed stories according to Mandler and Johnson's grammar and developed a prediction formula based on the tabulation of different types of basic nodes, weighting nodes according to their level in the underlying structure of the story. When they compared their results with the Fry formula, they found no relationship; as difficulty increased according to the Fry formula, their underlying structure or degree of difficulty remained the same across grade levels. They refined their formula and tested it by having subjects read silently and then retell the story. Analyses revealed no significant differences between recalls as a function of difficulty levels. Although this effort has resulted in a comparable measure of the difficulty of texts, at present it does not appear to predict appropriate placement of text materials in the same way readability formulas do.

The line of research related to the content structure of text has been very productive. It has led to a better understanding of the comprehension process and to the development of guidelines for text production (Armbruster, 1984). However, to date it has not led to a replacement for readability formulas.

The Cohesive Structure

In contrast to content structure, cohesive structure serves as the syntax of discourse. It is concerned with the interrelationships of ideas (Meyer & Rice, 1984). In effect, the cohesive structure is a roadmap to understanding. Although the reader may not understand the words specific to a particular field, the cohesive devices form the context in which words have meaning. They are "[the] mechanisms by which authors tie their materials together" (p. 325). While they are not the content to be learned, the cohesive structures help students order and organize new concepts.

Cohesion, as defined by Halliday and Hasan (1976), occurs when the interpretation of some element in the discourse is dependent upon the interpretation of another. This occurrence of a pair of related items forms a tie across the boundaries of sentences. It may be achieved through syntactical markers, such as conjunctions, or through semantic relationships, such as pronouns. This connection of terms across sentences represents a kind of "linguistic mortar" (Tierney & Mosenthal, 1982) that clearly defines the semantic continuity of a text.

Halliday and Hasan distinguish five classes of cohesive ties shown with simple examples in Figure 2.

Figure 2
Examples of Cohesive Ties

Class	Example
Reference	he, that, there
Substitution	one, same, do so
Ellipsis	
Conjunction	and, or, later
Lexical	kayak, boat

Their system of cohesion analysis provides multilevel information. It taps the commonly recognized cohesive devices and describes semantic networks as well. These devices generally operate within or between adjacent paragraphs – the microlevel. Through the coding of lexical items, chains mark the semantic continuity of any given passage across a larger level of organization. Regrettably, lexical analysis is the weakest point in the methodology because lexical ties are much more dependent on subjective judgment and prior experiential knowledge than the others. However, lexical ties are directly related to vocabulary knowledge and as such are a crucial element in analyzing comprehensibility. For example, preliminary research using this analysis system with college level textbooks indicates that substitution and ellipsis are rarely present in expository text, while lexical items predominate, accounting for more than 54 percent of all ties (Binkley, 1983).

Using Cohesion Analysis to Assess Readability

Cohesion analysis as described by Halliday and Hasan reduces text to counts of types of ties and distances. Work by Binkley and Chapman extends the applicability of cohesion analysis to assessments of readability.

Binkley and Chapman have developed a methodology that assesses the match between students and text materials intended for instruction. Their methodology looks at attributes of written text, comparing these attributes to students' development in much the way readability formulas were intended. However, it goes beyond current readability formulas.

- It accounts for more attributes of text than vocabulary difficulty and sentence length.
- It qualitatively evaluates reader performance.
- It provides diagnostic information on individual, small group, class, or school levels that could guide the planning of instruction or production of text materials.

The assessment process is intended for use when reading to learn is the objective. In such a setting there may be multiple goals for instruction, including learning how to learn from text and learning specific content. Consequently, it is essential that the text be within a manageable range for student readers. In addition, the teacher must know where and when a mismatch occurs between the author's assumed and the reader's actual ability so appropriate direct instruction may be provided. This assessment procedure pinpoints types of problems students are having. Teachers can then tailor instruction more appropriately.

There are essentially three stages to the assessment procedure: text analysis, design of a modified cloze procedure to reflect the attributes of the particular text, and administration and scoring of the cloze procedure.

What Cohesion Analysis Says about Specific Texts

Both similar and different attributes of text are assessed with readability formulae and with cohesion analysis. In the former, only two attributes are considered: vocabulary difficulty, which is measured against lists of familiar words or by counting the number of syllables, and sentence difficulty, which is measured by the average number of words per sentence.

In cohesion analysis more information is provided about vocabulary and syntactic complexity. The count of lexical items indicates the number of repetitions, the number of synonyms, the use of superordinate and subordinate terms, and the use of general classes of words. This type of count can be used to gather information pertaining to the ways children acquire word meanings, as well as their recognition of particular words. The count of conjunctive and reference items helps to assess the number and complexity of syntactical forms used in specific texts.

The pattern of ties that occurs across sample passages from a specific text reveals a great deal. For example, although each type of tie may be present in all written text, the distribution of ties differs among academic disciplines (Binkley, 1983). Science writings tend to repeat the same noun while social science writings depend more heavily on synonyms and superordinate and subordinate terms. Variation also occurs in the prevalence of types of conjunctions. These differences are neither good nor bad; they do, however, represent differences in argumentation style that may have implications for readers.

While there is evidence suggesting patterns of ties representative of each discipline (Binkley, 1983), there are even stronger indications that a specific pattern exists within a single textbook. This pattern might be called the signature or register of that book.

Binkley's method for determining a text signature is rooted in Halliday and Hasan's (1976) system for counting ties. Samples of text are randomly chosen from a textbook (a sample is defined as a unit of discourse that begins with a heading or subheading and ends at the next heading) and are analyzed with each tie and the distance between ties recorded. The distribution of types of ties and distances across the samples constitutes the signature or register of the textbook. It is this proportional variation and the specific subclasses of ties, as well as the number of ties that cohere over longer distances (i.e., more than five sentences) that differentiate one text from the other. As such, the register includes the specific content vocabulary, the relationship between ideas, the argument structure manifested by the use of reference and conjunctive ties, the syntax of the passages represented by the grammatical relationships of the ties, and syntactical markers of the macrostructure. Consequently, a register may be considered so distinctive as to constitute a separate genre. These factors, along with a reader's prior knowledge, influence the comprehensibility of a text.

The Instrument Design

Reading is an interaction between an author (who has made certain assumptions about an audience) and readers (who may or may not have the assumed attributes). Therefore, an assessment of a text separate from an assessment of the readers' characteristics cannot give a measure of the text's comprehensibility. In designing an assessment procedure, the emphasis should be on gathering information about text in relation to a particular body of students. To do so, the assessment instrument should relate the salient features of the text with the readers' ability to comprehend. The instrument will thus yield information about both the reader and text.

Measures of reading comprehension have taken many forms over the years. Cloze was one of the first techniques used to match students with appropriate reading materials. When used in schools as an instructional and testing tool for reading comprehension, cloze typically begins with an excerpt from which every *n*th word is deleted. Students are expected to fill in the resulting blanks by selecting a word. Since its introduction into education by Taylor (1953), much research has been conducted to test the cloze procedure, both as a device to measure comprehension and as a measure of readability (Beard, 1967; Bormuth, 1968; McKenna, 1978; Nesvold, 1972). More recently, research has focused on using cloze as an instructional technique (Jongsma, 1980; Kennedy & Weener, 1973). The cloze procedure has been shown to measure the difficulty of a text in a manner that is unlike readability formulas. Word and sentence length are not the variables considered. Instead, cloze measures a reader's response to linguistic variables, the language structure of text.

Shanahan, Kamil, and Tobin (1982) questioned the ability of cloze tests to measure the use of information across sentence boundaries. In their study, they administered three variations of cloze: standard cloze passages; the same passages with scrambled sentence sequences; and passages constructed by embedding single sentences from the original passages in other, nonsupportive text. They found no performance differences due to sentence order or to the presence of supportive text. Therefore, they concluded the cloze procedure might be limited in measuring the integration of intersentential information. They do suggest that "It might be possible to design cloze tests to measure this ability."

In contrast to a standard cloze procedure, the deletions in Binkley and Chapman's assessment instrument are based on the signature or register of the textbook under consideration. Therefore, the assessment instrument tests the language demands of the textbook going beyond intra to intersentential information integration. When coupled with the grading system, the procedure allows for qualitative descriptions of student performance.

Three criteria are considered in making deletions. First, because the distribution of ties in the textbook marks the syntax prevalent in the text, deletions are made to reflect the proportion of ties in the textbooks. For example, if pronouns are used 20 percent of the time throughout the text, 20 percent of the deletions in the modified cloze procedure should be pronouns. In this manner, the assessment instrument highlights the language structures that set this writing apart from other writings, because it marks the relationships between ideas.

Second, consideration is given to the pattern of distance between ties. Ties are generally between adjacent sentences. They occur less often between sentences two to three sentences apart. Rarely do they occur between sentences greater than five sentences apart. Ties in adjacent sentences reflect micro or local coherence. In contrast, ties across paragraph boundaries tend to reflect macro or global coherence. Therefore, deletions should represent the proportion of various distances. This dimension gives information about the reader's use and understanding of the micro and macrostructures of particular texts.

Finally, consideration is given to making deletions that relate to and trace the major chains central to a particular excerpt. Here the assessment procedure relates closely to the lexical chains and the semantic knowledge necessary for understanding specific content. This, too, is an essential element of the macrostructure.

Scoring Student Responses

The Binkley and Chapman instrument is administered in the same manner as a standard cloze procedure. However, the scoring system is significantly different. In a standard cloze procedure, responses are right if the replacement word is the exact word deleted. The student's score is the number of responses that are the same as the original. Based on this number, a book is determined to be at the independent, instructional, or frustration level for a particular student. No diagnostic information is provided as to type of errors or difficulty.

In the Binkley and Chapman system, student responses to the cloze procedure are placed on a continuum from inappropriate (i.e., no relation to the materials) to syntactically correct to syntactically and semantically correct (Chapman, 1979a, b, c; 1980; 1983a, b). The initial analysis of student responses yields a frequency of response for each deletion. All responses are reported with a count of how many students chose each response. The responses are recorded so they are positioned along a continuum. The criteria for

Figure 3
Summary of Criteria for Allocating Responses

Position 1 (P1) Prereading
- omissions
- unrecognizable responses
- response from VPF
- response unacceptable in one clause element

P1 Transition (response is partially acceptable)
- achieved by ignoring other words in clause element
- achieved by overrunning punctuation and combining with word(s) from following clause elements

Position 2 (P2) Beginning Reading (clause structure perceived)
- word complex responses – acceptable in one clause element only (i.e., all other contexts are ignored)
- group complex responses – acceptable in clause (complex) but lacking evidence of cohesion and register
- clause complex responses – acceptable in clause (complex) but lacking evidence of cohesion and register

P2 Transition
- response shows evidence of cohesion and appropriate register but contains errors in lexogrammatical structure(s)

Position 3 (P3) Developing Reading Fluency
- responses indicate that clause structure perceived; cohesion perceived but achieved differently from author; possible errors of field mode and tenor
- responses indicate that clause structure perceived; register is appropriate but cohesion achieved differently from author

P3 Transition
- structure perceived, register appropriate, cohesion perceived but not author's word

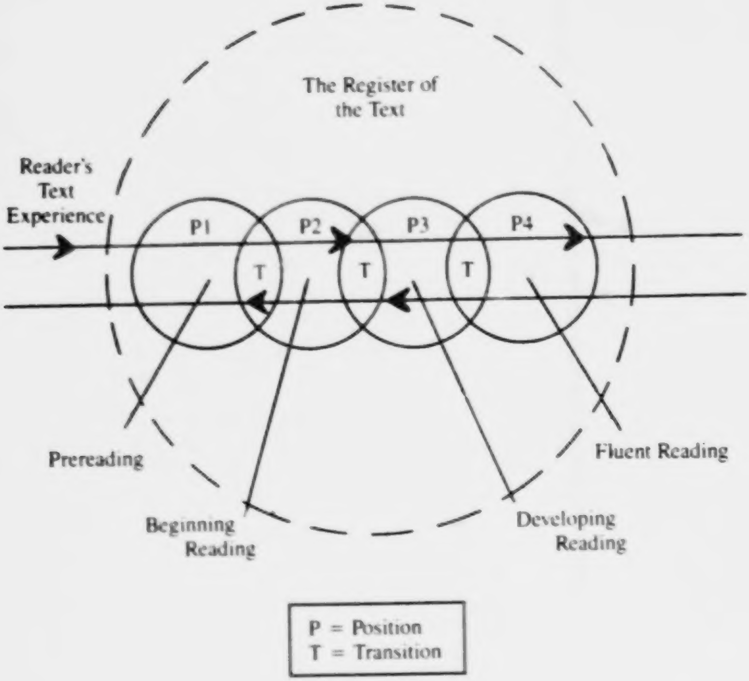
Position 4 (P4) Fluent Reading
- criterion met so that either author's or teacher's word is provided

Reprinted by permission of L.J. Chapman, 1983.

assigning a response to a position, which were established by Chapman (1983c), are summarized in Figure 3. Based on this analysis of student responses, the reader's abilities may be characterized on the reading development continuum (pictured in Figure 4) developed by Chapman (1983c).

Qualitative analyses of the types of errors make possible an assessment of the problems students may have with a particular text, i.e., whether their misunderstandings are based on vocabulary/se-

Figure 4
The Reading Development Continuum


The Register of the Text
Reader's Text Experience
P1
P2
P3
P4
T
T
T
Prereading
Beginning Reading
Developing Reading
Fluent Reading
P = Position
T = Transition

Reprinted by permission of L. J. Chapman, 1983.

mantics, syntactical/language structures, or organization. Consequently, teachers have more information for determining appropriate instruction.

The coding described allows for interpretation of responses to individual deletions, each sample, or across samples for individuals or class groups. In this manner, cohesion analysis results in a measure of readability that provides more diagnostic information than classic readability formulas and can account for intersentential integration of information.

An Example of the Use of the Binkley and Chapman Procedure

The Binkley and Chapman procedure was recently tested on a new fourth grade social studies textbook. The following discussion of that pilot test will clarify use of the procedure.

For the purpose of the pilot test, five random samples from the textbook were analyzed so that a record of each tie and its distance was recorded in the method prescribed by Halliday and Hasan (1976). The counts were then summarized so that distribution of ties, in their broadest categories, could be assessed. This summary of the distribution of ties appears in Figure 5.

Using a chi-square test, we determined that the distribution of ties was homogeneous across the samples. Sample B represented the distribution most closely and was selected as the excerpt to be used for the pilot test.

As described, deletions were made from the sample so that seven (27 percent) deletions were reference items, two (7 percent) were conjunctions, and sixteen (63 percent) were lexical. Within each class of ties, specific deletions depended upon the number within each subclass of ties and the number of ties at a given distance. For example, in the case of lexical ties, the distribution reflected the percentage of ties using the same item, general term, and collocation. The result was a cloze procedure with twenty-five deletions.

Figure 5
The Distribution of Ties across Samples
Fourth Grade Social Studies Textbook

Sample	A	B	C	D	E	Total	Percent
Reference	41	30	25	14	32	142	27
Substitution	1	–	2	1	1	5	1
Ellipsis	1	2	2	–	2	7	1
Conjunction	4	9	9	3	10	35	7
Lexical	73	69	80	44	60	326	63
Total	120	110	118	62	105	515	

Figure 6
Distribution of Responses along Reading Development Continuum

P1 Prereading	14 percent
P2 Beginning Reading	17 percent
P3 Developing Reading	42 percent
P4 Fluent Reading	28 percent

Percentages represent the number of responses in each category compared with a total of 1,325 responses.

The Results

Fifty-three fourth grade students were tested. Their responses were scored as outlined.

Based on analysis of the distribution of responses (fifty-three responses to each of the twenty-five items) along the reading development continuum (as outlined in Figure 6), we conclude that this text is well matched with this group of students. Less than one-third of the student responses were below the developing reading level. Six of the subjects were non-English speaking students who accounted for a large proportion of omissions and prereading responses.

Examining the frequency of responses, we determined that the fifty-three students had little difficulty with reference cohesion types in this text. This is evident by the frequency of correct responses to deletions requiring pronouns (Figure 7).

The omission rate is notably low. Where there were high numbers of omissions, as in items 4 and 25, we believe the result was due to the newness of the form for fourth graders. Both these items required students to use *also*.

> They named the river the James, in honor of their king, James I. They – 4 – named their settlement Jamestown in his honor.
>
> By 1733, Jamestown was only one of many English settlements in Virginia. The English had – 25 – started three new colonies south of Virginia – North Carolina, South Carolina, and Georgia.

Figure 7
Frequency of Student Responses

Author's word	Five Most Frequent Answers					Omissions
	First	Second	Third	Fourth	Fifth	
1. came	wanted N = 19 P = 3	decided N = 6 P = 3	had N = 5 P = 3	came N = 3 P = 4	started N = 3 P = 3	1
2. the	some N = 3 P = 3	those N = 3 P = 3	English N = 2 P = 2t	many/few N = 2 P = 3	early N = 1 P = 2t	3
3. they	they N = 35 P = 4	settlers N = 2 P = 2t	Indians N = 2 P = 2	was N = 2 P = 2	Jamestown N = 1 P = 2	5
4. also	had N = 14 P = 3	also N = 8 P = 4	all N = 5 P = 3	finally N = 2 P = 3t	people N = 2 P = 2t	12
5. settlement	town N = 18 P = 3	settlement N = 5 P = 4	king N = 5 P = 1	country N = 4 P = 3	city/land N = 3 P = 3	5
6. colonists	people N = 20 P = 3	settlers N = 8 P = 3t	king N = 4 P = 2	settlement N = 3 P = 2t	group/men N = 2 P = 3	2
7. they	they N = 44 P = 4	most N = 1 P = 3	many N = 1 P = 3	James N = 1 P = 3	nobody N = 1 P = 3	3
8. colonists	people N = 26 P = 3	settlers N = 12 P = 3t	explorers N = 4 P = 3	men N = 4 P = 3	villagers N = 1 P = 3	2
9. England	England N = 16 P = 4	Jamestown N = 9 P = 2	rest N = 3 P = 3	relax N = 2 P = 3	sleep N = 2 P = 3	5
10. stayed	lived N = 22 P = 3	stayed N = 15 P = 4	worked N = 2 P = 3	were N = 3 P = 2t	belonged/ remained N = 2 P = 3	5
11. arrived	came N = 38 P = 3	arrived N = 6 P = 4	John N = 2 P = 2	began N = 1 P = 2	called N = 1 P = 2	3
12. his	his N = 43 P = 4	the N = 5 P = 3t	who N = 1 P = 2t	was N = 1 P = 1t	Smith N = 1 P = 1	2
13. crop	way N = 18 P = 2	crop N = 5 P = 4	farm N = 4 P = 2	animal N = 3 P = 2	place/tobacco N = 2 P = 3	6
14. Rolfe	he N = 15 P = 3	Rolfe N = 10 P = 4	they N = 8 P = 2t	John N = 6 P = 3	John Rolfe N = 5 P = 4	5
15. Caribbean	the N = 33 P = 3	his N = 3 P = 3	Caribbean N = 2 P = 4	some N = 2 P = 3	then/there N = 1 P = 2t	2

Figure 7 (continued)
Frequency of Student Responses

	Five Most Frequent Answers					
Author's word	First	Second	Third	Fourth	Fifth	Omissions
16. grow	plant N = 10 P = 3	make N = 9 P = 3	raise N = 7 P = 3t	grow N = 5 P = 4	get N = 4 P = 3	3
17. sell	sell N = 21 P = 4	make N = 7 P = 2t	raise N = 4 P = 2t	grow N = 3 P = 2t	use/plant N = 2 P = 3	3
18. their	the N = 14 P = 3t	their N = 13 P = 4	all N = 3 P = 3t	and N = 3 P = 3	now/then/later N = 1 P = 3	8
19. tobacco	tobacco N = 20 P = 4	they N = 16 P = 2t	he N = 3 P = 2t	John Rolfe N = 2 P = 2t	it N = 2 P = 3	7
20. gold	anything N = 7 P = 2t	gold N = 5 P = 4	tobacco N = 3 P = 2	money N = 3 P = 2t	much N = 3 P = 2	10
21. river	great N = 4 P = 3	high N = 4 P = 3	deep N = 2 P = 3	low N = 2 P = 3	Jamestown N = 2 P = 3	9
22. new	new N = 8 P = 4	the N = 6 P = 3	early N = 4 P = 3	people N = 3 P = 2t	years N = 3 P = 2	7
23. coast	Jamestown N = 5 P = 2t	Caribbean N = 5 P = 3	east N = 4 P = 3t	settlers N = 3 P = 2t	beginning N = 3 P = 2t	8
24. Jamestown	there N = 13 P = 2	Jamestown N = 10 P = 4	tobacco N = 8 P = 2	it N = 4 P = 3	John Rolfe N = 2 P = 2	10
25. also	already N = 14 P = 3	now N = 10 P = 3	only N = 4 P = 3	also N = 1 P = 4	settlers/people N = 1 P = 2	10

(N) Number of students giving this response

(P) Position on reading continuum where
1 = Prereading
2 = Beginning Reading
3 = Developing Reading
4 = Fluent Reading
5 = Transition

(See Figure 4 for further description.)

Items 20 and 24 also had high omission rates. These two items were comparatively distant from the referent. For example, item 20

> The people of Jamestown now understood that the great resource of their new home was not – 20 – but farm land.

called for *gold* as the response. This was directly related to the first sentence of the second paragraph, seventeen sentences away, or to the notion of gold as a form of cash. Similarly, in item 24

> By 1733, – 24 – was only one of many English settlements in Virginia

students were required to jump back in the macrostructure to fill in Jamestown although the topic of the preceding paragraph had been the land resources.

Students had the greatest difficulty with register specific words, i.e., colonist and settlement. While their responses were usually in line with the concept, they did not understand the size differential between a colony and a city.

In summary, results indicated that the book was appropriate for the instruction of most of the students. The qualitative analysis of responses to particular items suggested possible teaching strategies (i.e., how the teacher might wish to introduce new vocabulary), or raised questions about possible revisions to the textbook before publication (i.e., whether *also* is appropriate for fourth graders).

Potential Applications of This Methodology

The methodology has important potential. Teachers could use the information from the qualitative analysis as a guideline for lesson planning. Too often, teachers who believe students do not understand reteach and drill the lesson in the same manner in which it was originally presented. As documented, teachers would be able to discriminate between types of errors and might develop alternative strategies.

The methodology would be very helpful to adoption committees within a school or district. Through the use of sample field tests, it would be possible to rank order the textbooks under consideration. Because of the well documented limitations of readability formulas and the constraints of the adoption process, this methodology would serve as a more objective measure of suitability.

The methodology could be of use to publishers as well. If they used the methodology during the development phases of new textbooks, they would have a measure of comprehensibility beyond readability formulas. Based on the qualitative analyses, they might consider rewriting parts of books to avoid difficulties students might have. Information about specific student needs could be included in a teachers' manual.

This pilot test demonstrates the potential of the methodology. Chapman is building a data base of student responses that may eliminate the need for field testing and could drastically reduce the labor intensity of the process.

Conclusions

There are several significant points to stress. First, classic readability formulas serve an important purpose. They are intended to and do predict an approximate level of difficulty. Critics would like these formulas to account for more of the complexities of text. Readability research has shown that the addition of attributes does not increase the reliability of the formulas.

Critics claim readability formulas are detrimental to textbook production. This is true when the formulas are applied in ways that were never intended, but this is not the fault of the formulas. Other sources of information about the quality of texts are available.

Even though it has roots in the works of Aristotle, text research is a comparatively new phenomenon. Systematic study relating text features to learning dates from the late sixties. Textlinguistics, which looks at discourse beyond the sentence, dates from the late seventies. At this time it is unrealistic to expect an elegant formula to objectively measure text difficulty in the ways readability formulas do.

New methods of assessing text have proven productive. They have provided researchers with information about the reading process in general and its relationship to attributes of text. Guidelines for text production are an important step toward improving texts. As more researchers address the issue, new ways of assessing text may evolve, yielding simple, objective ways of calibrating text.

References

Armbruster, B. The problem of "inconsiderate text." In G. Duffy, L. Roehler, and J. Mason (Eds.), *Comprehension instruction: Perspectives and suggestions*. New York: Longman, 1984.

Bartlett, B.J. *Top level structure as an organizational strategy for recall of classroom text.* Unpublished doctoral dissertation, Arizona State University, 1978.

Bartlett, F.C. *Remembering*. Cambridge: Cambridge University Press, 1932.

Beard, J.G. Comprehensibility of high school textbooks: Association with content area. *Journal of Reading*, 1967, *11*, 229-234.

Binkley, M.R. *A study of cohesive structure in text materials for differing academic disciplines*. Unpublished dissertation, George Washington University, 1983.

Black, J.B., and Wilensky, R. An evaluation of story grammars. *Cognitive Science*, 1979, *3*, 213-230.

Bormuth, J.R. Cloze test readability: Criterion referenced scores. *Journal of Educational Measurement*, 1968, *5*, 189-196.

Bower, G.H. Experiments on story understanding and recall. *Quarterly Journal of Experimental Psychology*, 1976, *28*, 511-534.

Britton, B.K., Meyer, B.J., Simpson, R., Holdredge, T.S., and Curry, C. Effects of the organization on memory: Tests of two implications of a selective attention hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, *5*, 496-506.

Chapman, L.J. *A reading development continuum: A framework for teaching reading?* Paper presented to the Ninth Conference of the Australian Reading Association, Tasmania, 1983c.

Chapman, L.J. *A study in reading development: A comparison of the ability of 8, 10, and 13 year old children to perceive cohesion in their school texts.* Paper presented at the Nineteenth Annual Conference of the United Kingdom Reading Association, 1983a.

Chapman, L.J. Confirming children's use of cohesive ties in text. *The Reading Teacher*, 1979b, *33*, 317-322.

Chapman, L.J. *Reading development and cohesion*. London: Heinemann, 1983b.

Chapman, L.J. Some developments in textlinguistics: Implications for the teaching of reading. 1979c. (ED 174 767)

Chapman, L.J. The perception of language cohesion during fluent reading. In P. Kolers, M. Wrolstad, and H. Bouma (Eds)., *Processing of visible language, volume 1*. New York: Plenum, 1979a.

Chapman, L.J. *The reader and the text*. London: Heinemann, 1980.

Charniak, E. Towards a model of children's story comprehension. Cambridge, MA: MIT AI Lab, Memo TR-266, 1972.

D'Angelo, F.J. *A conceptual theory of rhetoric*. Cambridge, MA: Winthrop, 1975.

Dawes, R.M. Memory and distortion of meaningful written material. *British Journal of Psychology*, 1966, *57*, 77-86.

Duchastel, P.C. Learning objectives and organization of prose. *Journal of Educational Psychology*, 1979, *71*, 100-106.

Fillmore, C. The case for case. In E. Bach and R.T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart & Winston, 1968.

Flower, L.S., and Hayes, J.R. Problem solving strategies and the writing process. *College English*, 1977, *39*, 449-461.

Frase, L.T. Integration of written text. *Journal of Educational Psychology*, 1973, *65*, 252-261.

Frederiksen, C.H. Acquisition of semantic information from discourse: Effects of repeated exposures. *Journal of Verbal Learning and Verbal Behavior*, 1975a, *14*, 158-169.

Frederiksen, C.H. Effects of task induced cognitive operations on comprehension and memory processes. In R.O. Freedle and J.B. Carroll (Eds.), *Language comprehension and the acquisition of knowledge*. New York: Wiley, 1972.

Frederiksen, C.H. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology*, 1975b, *7*, 371-457.

Grimes, J.E. *The thread of discourse*. The Hague: Mouton, 1975.

Halliday, M.A., and Hasan, R. *Cohesion in English*. London: Longman, 1976.

Haring, M.J., and Fry, M.A. Effect of pictures on children's comprehension of written text. *Educational Communication and Technology*, 1979, *27*, 185-190.

Hasan, R. *Coherence and cohesive harmony*. Paper presented at the Annual Convention of the International Reading Association, May 1980.

Jongsma, E.A. *Cloze instruction: A second look*. Newark, DE: International Reading Association, 1980.

Kennedy, D.K., and Weener, P. Visual and auditory training with the cloze procedure to improve reading and listening comprehension. *Reading Research Quarterly*, 1973, *8*, 524-541.

Kintsch, W. *Memory and cognition*. New York: Wiley, 1977.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum, 1974.

Kintsch, W., and van Dijk, T.A. Recalling and summarizing stories. *Languages*, 1975, *40*, 96-116.

Mandler, M., and M.S. Johnson. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111-151.

Marshall, N., and Glock, M.D. Comprehension of connected discourse: A study into the relationship between the structure of text and information recalled. *Reading Research Quarterly*, 1978, *14*, 10-56.

McKenna, M.C. Diagnostic uses of the cloze procedure. *Selected articles on the teaching of reading*, set B, No. 41. New York: Barnell Loft, 1978.

Meyer, B.J. Basic research on prose comprehension: A critical review. In D.F. Fisher and C.W. Peters (Eds.), *Comprehension and the competent reader: Interspecialty perspectives*. New York: Praeger, 1981.

Meyer, B.J. *Idea units recalled from prose in relation to their position in the logical structure, importance, stability, and order in the passage*. Unpublished master's thesis, Cornell University, 1971.

Meyer, B.J. Organizational patterns in prose and their use in reading. In M.L. Kamil and A.J. Moe (Eds.), *Reading research: Studies and applications*. Clemson, SC: National Reading Conference, 1979, 109-117.

Meyer, B.J. *The organization of prose and its effects on memory*. Amsterdam: North Holland Publishing, 1975.

Meyer, B.J. The structure of prose: Effects on learning and memory and implications for educational practice. In R.C. Anderson, R.J. Spiro, and W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum, 1977, 179-200.

Meyer, B.J., and Freedle, R.O. Effects of discourse type on recall. *American Educational Research Journal*, 1984, *21* (1), 121-143.

Meyer, B.J., Haring, M.J., Brandt, D.M., and Walker, C.M. Comprehension of stories and expository text. *Poetics*, 1980, 9, 203-211.

Meyer, B.J., and Rice, G.E. The structure of text. In P.D. Pearson (Ed.), *Handbook of reading research*. New York: Longman, 1984.

Nesvold, K.J. Cloze procedure correlation with perceived readability. *Journalism Quarterly*, 1972, *49*, 592-594.

Rice, G.E. On cultural schemata. *American Ethnologist*, 1980, *1*, 152-171.

Rice, G.E. *The role of cultural schemata in narrative comprehension*. Social Sciences Research Report No. 10. Irvine, CA: University of California, 1978.

Rumelhart, D.E. Notes on a schema for stories. In D.G. Bobrow and A.M. Collins (Eds.), *Representation and understanding*. New York: Academic Press, 1975, 211-236.

Rumelhart, D.E. Understanding and summarizing brief stories. In D. LaBerge and S.J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, NJ: Erlbaum 1977, 265-304.

Schank, R.C., and Abelson, R.P. *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum, 1977.

Shanahan, T., Kamil, M.L., and Tobin, A.W. Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*, 1982, *27* (2), 229-255.

Simmons, R.F. *Towards a computational theory of textual discourse*. Austin: Department of Computer Science, University of Texas, 1978.

Stein, N.L., and Glenn, C.G. An analysis of story comprehension in elementary school children. In R.O. Freedle (Ed.), *New directions in discourse processing*. Norwood, NJ: Ablex, 1979.

Swanson, C. *The effects of readability and top level structure on ninth graders' recall of textual materials*. Unpublished doctoral dissertation, Arizona State University, 1979.

Taylor, W.L. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, *30*, 415-433.

Templeton, S., and Mowery, S. Readability, basal readers, and story grammar: What lies beneath the surface? *Reading World*, March 1985, 40-47.

Thorndyke, P.W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, 9, 77-110.

Tierney, R.J., and Mosenthal, J. *Discourse comprehension and production: Analyzing text structure and cohesion*. Technical Report No. 152. Urbana, IL: University of Illinois, 1982.

van Dijk, T.A. Relevance assignment in discourse comprehension. *Discourse Processes*, 1979, *2*, 113-126.

van Dijk, T.A. Semantic macrostructures and knowledge frames in discourse comprehension. In M. Just and P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Erlbaum, 1977.

Walker, C.H., and Meyer, B.J. Integrating different types of information in text. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 263-275.

Wilensky, R. Why John married Mary: Understanding stories involving recurring goals. *Cognitive Science*, 1978, *2*, 235-266.

Beverley L. Zakaluk
S. Jay Samuels

7

Toward a New Approach to Predicting Text Comprehensibility

Psychological research on what makes a text readable has a relatively brief history, going back only about fifty years. Concern about text comprehensibility, however, can be traced back 2,500 years to Greek scholars who were attempting to train Athenian lawyers in the arts of policy analysis, exposition, and persuasion – topics that constitute the roots of classical rhetoric. While recognizing the important work on readability carried out by rhetoricians, this chapter nevertheless has a psychological orientation.

For reading educators, perhaps the most important new understanding about readability has come about because of a shift in emphasis in psychological study. Behaviorism has been abandoned in favor of a cognitive approach to human information processing. Researchers are increasingly aware that whenever the reading level of the material changes, the nature of cognitive processing changes also. Both decoding ability and text topic familiarity influence reading comprehension performance. When a selection is estimated to be at the third grade readability level, we assume it is easy to comprehend. If, however, the reader is at the beginning stages of reading and is neither accurate nor automatic at decoding, comprehension will be low (Perfetti & Lesgold, 1977, 1979; Samuels, 1977). Similarly, a skilled reader will experience difficulty comprehending even relatively simple text when the topic is completely unfamiliar (Kintsch & Miller, 1984; Kintsch & van Dijk, 1978; Kintsch & Vipond, 1979). To their detriment, readability formulas in use today concentrate only on text characteristics, totally neglecting how cognitive processing factors influence the comprehensibility of text.

We begin this chapter by describing the use of readability formulas in matching readers with material appropriate to their instructional level. Then, in keeping with an interactive view of the reading process, we review outside and inside the head factors that influence comprehensibility. Finally, we suggest a new way to predict text comprehensibility.

Matching Readers with Appropriate Materials

An important instructional mandate is to assign written material at levels corresponding to individual reading achievement. In elementary school classrooms, however, reading ability may vary from three to seven or eight grade levels. Further, as students move upward through the grades the range increases; the higher the grade level, the greater the spread of classroom reading achievement within each class (Balow, 1962; Betts, 1957; Bond & Wagner, 1966). Yet to teach reading effectively or to help students gain information from text requires a match between the difficulty of the reading material and the reading ability of the child since students make optimal gains when instructed at a level where they can succeed (Dunkeld, 1970; Johnson, Kress, & Pikulski, 1986; Scarborough, Bruns, & Frazier, 1957). Gray and Leary (1935, p. 5) suggest that "to get the right book into the hands of the right reader" is a pressing responsibility.

To solve the problem of matching readers with appropriate material, researchers developed prediction formulas for estimating the difficulty of books and reading selections. Typically, two phases are involved in developing a readability formula with which to judge text difficulty (Bormuth, 1971; Dale & Tyler, 1934; Gray & Leary, 1935; Pearson, 1974). In the first phase, the developer selects samples of reading materials at successive levels, then constructs questions to test readers' ability to comprehend each passage. Next, the developer chooses a target population to read the information and to complete the corresponding test items. Subsequently, the developer uses the resultant mean scores to establish an index of difficulty for the material. This index serves as a criterion measure to be used as a dependent variable in the second part of the study.

In the second phase, the developer quantifies factors believed to be predictive of reading difficulty. These factors include such semantic elements as the specialized technical vocabulary in a passage; the easy words; and the hard, nontechnical words. Syntactic elements considered are the type and length of sentences and the number of clauses and prepositional phrases. The developer establishes correlations with the previously identified index of difficulty, then determines which elements relate most highly to the criterion measure. Factors failing to improve the predictive power of the equation are dropped, and a final multiple regression equation is developed. The resulting formula is used to estimate the reading difficulty of a wide variety of printed information.

In general, readability formulas are based on two factors: word difficulty (as measured by familiarity, frequency, or length) and sentence complexity. These two variables represent the highest loadings on the regression equations used to predict text difficulty. In most cases, the result is that these two outside the head text variables alone are used almost exclusively to judge reading ease. Critics of readability formulas most often cite the formulas' reliability, criterion validity, and disregard for higher level text organization. Using formulas as prescriptions for writing also has been censured. (See earlier chapters in this volume.) Elaborations of these issues follow.

Outside the Head Text Factors

Limitations of Existing Readability Formulas

Interformula reliability. A critical test of any formula is to compare readability estimates on the same passage, applying the formula under consideration and other formulas of established credibility. Studies indicate that readability levels differ depending on which formula is used. That is, the readability level of a prose selection might be rated as most difficult by one readability formula and least difficult by another formula. McConnell (1982), for example, found discrepant results for nine college level introductory economics texts. One text had a grade level equivalent of 11.1 when readability was calculated with the Dale-Chall formula, 8.2 when the Modified Dale-Chall formula was applied, but 10.7 when readability was estimated by the Fry formula. Even more striking is the fact that rank order of difficulty for the set of books changed according to the formula employed. For example, the difficulty level of one economics text was rated by one formula as the easiest of the nine texts surveyed and by another formula as the next to the hardest.

Criterion validity. A problem with criterion validity for the Lorge, Flesch, and Dale-Chall formulas can be traced back to their origin. For convenience (and perhaps because they believed the passages to be standardized) these investigators administered the previously graded McCall-Crabbs Standard Test Lessons as the reading selections on which to base the criterion index of difficulty. These researchers correlated text variables, such as word length and sentence complexity, with McCall-Crabbs passage grade equivalent scores, choosing a grade score that designated a comprehension performance of either 50 or 75 percent. The McCall-Crabbs passages, however, are inadequately normed and were never intended to be employed as criteria for readability formulas. As Stevens (1980a) explains, only pupils from New York City public schools were used in standardizing the selections. The grade scores for each McCall-Crabbs passage were obtained by smoothing the curve on a graph connecting student performance on the Thorndike-McCall Reading Scales with their test lesson performance. Since the grade score equivalents were designed to serve as approximations, records of their derivation were never kept. Such casual and insufficient norming suggests McCall-Crabbs grade scores are unreliable and invalid criteria for developing readability measures.

Disregard for higher level text organization. A basic limitation of readability formulas is that they ignore such critical text factors as cohesiveness and macrolevel organization. Thus, it is possible to randomize every sentence in a text without changing the tabulated readability. The assigned readability level of an eight word sentence in jumbled order would not differ from the assigned readability level of an eight word sentence in normal order (Marshall, 1979). The well organized and the poorly organized text would have the same designated difficulty level.

Prescription for writing. Another drawback of readability formulas, resulting in a disservice to students, is the recommendation by formula authors that readability indexes be used as guides for composing more comprehensible text (Dale & Chall, 1948; Dale & Tyler, 1934; Flesch, 1948; Gray & Leary, 1935; Lorge, 1944). Gray and Leary suggested shortening the average sentence length and decreasing the number of prepositional phrases to increase reading ease. Flesch proposed that formula study be part of the curriculum in composition, creative writing, journalism, and advertising courses. In fact, rewriting text to conform to a prescribed reading level may result in text that is more difficult to read. Shorter sentences may not be the answer.

Coleman (1971) pointed out that if the number of ideas is held constant, understanding is enhanced when text is elaborated on or paraphrased. Pearson (1974) contended that text is easier to comprehend when ideas are stretched over several clauses instead of packaged into a single clause. Grammatical complexity may add to text comprehensibility.

The following illustration, in which the first sentence is easier to understand than its two sentence counterpart, reinforces this point.

1. People thought dew fell from the sky because it can be seen only in the morning.
2. People thought dew fell from the sky. It can be seen only in the morning.

In the first sentence, the causal relationship remains intact. By stating the ideas in two sentences, the author is trying to decrease reading difficulty by simplifying grammatical complexity (omitting the word *because*) and by reducing sentence length. Thus the reader must make an inference between the two statements. Tailoring text to conform to the constraints of readability formulas may detract from, rather than enhance, text comprehensibility.

As shown in the next example, "argument repetition" (Kintsch & van Dijk, 1978; Kintsch & Vipond, 1979), or the reiteration of key words and concepts, also enhances the logical flow of meaning from one sentence to the next.

> John likes walking to the new shopping mall. The shopping mall has many quaint shops.

Kintsch and his colleagues (1975) found that for university students the recall of words and concepts increased as a function of the number of times the argument or word concept was repeated in the text. Similarly, Manelis and Yekovich (1976) demonstrated that despite longer and more complicated sentences, repetition of the same concepts across sentences facilitated recall. When arguments are repeated, integrating the ideas in the text and making the relationship between the author's ideas explicit, the text becomes easier to process and hence more readable. The directive to simplify and increase reading ease through shortening is thus, paradoxically, ill advised because of the extra processing burden imposed.

A more global problem exists in regard to rewriting stories to make them more readable. Many folktales, fables, and myths that appear as basal reader selections have been intentionally altered to keep sentences short and to employ high frequency words. Nonetheless, each word change or deletion can result in distortion of semantic content and syntactic flow. If causal relationships are omitted, misunderstanding, not better comprehension, may be the outcome. Both common sense and literary taste should mitigate against the practice of tampering with sentence length and vocabulary load. As observed by Finn (1975), colorful low frequency words may carry nuances of meaning lacking in familiar words. Rare words also may be repeated within a text, thereby increasing reading ease. In the case of informational text, mastering technical words may be essential to understanding the substance of the material (Nelson-Herber, 1985).

Adjunct Comprehension Aids

Regardless of criticism concerning the limitations of readability formulas, the formulas are products of years of cumulative research originating in the twenties (Dale & Chall, 1948; Dale & Tyler, 1934; Flesch, 1948; Gray & Leary, 1935; Lively & Pressey, 1923; Lorge, 1939; Vogel & Washburne, 1928). Among the variables studied, word difficulty and sentence length have been consistently identified as factors that differentiate text difficulty levels. Making a judgment about the readability of respective texts on the basis of these two factors may be a relatively efficient approach to estimating comprehensibility.

Further, with a group of less competent readers and a mandated text, educators may facilitate comprehension by highlighting important information in the text. Such modifications are called adjunct comprehension aids.

Among adjunct aids, instructional objectives and questions placed within the text itself are relatively simple techniques that teachers can easily adopt to enhance text comprehensibility.

Interspersed questions. Empirical evidence supports the general claim that the inclusion of questions in text facilitates learning (Anderson, 1980; Anderson & Biddle, 1975; Faw & Waller, 1976). Research has shown that questions placed after the text enhance learning, both when followup questions are identical to the criterion questions and when they are new items. On the other hand, adjunct questions appearing before the text are effective only when the followup questions are similar (Frase, 1967; Rothkopf, 1966). Questions interspersed throughout the text that appear in close proximity to the information on which they are based also are facilitative (Rothkopf & Bisbicos, 1967). In addition, higher order questions that require students to answer beyond the level of literal response are equally appropriate, both when followup questions are identical to the original questions and when questions are new (Andre, 1979). Depending on both the type and sequencing of questions, their placement in expository text improves learning and makes text more readable for college students.

Instructional objectives. Instructional objectives are statements or study goals presented at the beginning of a text that suggest what the reader should know after reading. Research findings are inconclusive on the effectiveness of placing instructional objectives within the textual format as an aid to learning (Anderson, 1980; Duchastel & Merrill, 1973; Faw & Waller, 1976; Jenkins & Deno, 1971). For example, specific objectives facilitate intentional learning for high school and college age students, but have a deleterious effect on incidental learning if passages are lengthy (Kaplan, 1974; Rothkopf & Kaplan, 1972). Anderson's overall conclusion seems justified; learning is found to be greater when instructional objectives are explicitly stated than when objectives are not provided. A key point is noteworthy here. Objectives tend to enhance learning only when they direct students to focus on information they would not otherwise perceive as important (Duell, 1974). Too many objectives may be overwhelming, and students may disregard them altogether (Duchastel & Merrill, 1973).

Theoretically, the use of objectives in informational text should help students identify information that is important to remember. Rothkopf (1966) advocates the use of such instructional objectives because they do not detract from learning and, in particular cases, may increase learning from text. While this conclusion is logical, further study is required to see whether such research may be generalized and instructional objectives used to increase the comprehension performance of elementary school pupils.

Readability is not an inherent property of text, but the result of an interaction between a set of particular text characteristics and the information processing characteristics of individual readers (Kintsch & Vipond, 1979). Text factors alone cannot determine readability; readers' prior knowledge and understanding influence comprehensibility and recall.

Inside the Head Cognitive Factors

Background Knowledge

World knowledge plays an important role in reading comprehension (Adams & Collins, 1979; Armbruster & Anderson, 1981; Ausubel, 1960; Rumelhart & Ortony, 1977). For example, Kintsch and his colleagues (1975) reported a large difference in reading times for college students when paragraphs were easy and dealt with well known topics from classical history. When paragraphs were more demanding, and focused on scientific topics about which subjects possessed little or no previous knowledge, students took longer to read the text and recalled less information. Chiesi, Spilich, and

Voss (1979) also found that it was easier for college age students to learn more about a particular topic when they possessed high prior knowledge of that topic.

Research findings are similar for subjects at other age and grade levels. Pearson, Hansen, and Gordon (1979), working with second grade children, and Stevens (1980b), working with ninth graders, showed that subjects highly familiar with the topic not only had better comprehension, but recalled more than subjects not highly familiar with the topic. Taylor (1979) found that poor fifth grade readers' comprehension was diminished when their prior knowledge was low, but the poor readers were able to comprehend adequately when topics were familiar and written at appropriate difficulty levels. Dooling and Lachman (1971), Bransford and Johnson (1972), and Bransford and McCarrell (1974) demonstrated that until background knowledge is brought to a text, the text may seem incomprehensible. Well written texts signal to the reader what background knowledge must be activated to enhance processing.

Insight into the role that prior knowledge plays in facilitating text processing has been obtained from such studies as that of Blachowicz (1977), who found that readers in second, fifth, and seventh grades and college level adults frequently remembered more than what was contained in the sentences read as evidenced by the false identification of inference statements as statements they had encountered in the passages. Blachowicz hypothesized that during reading and recall readers use their knowledge of the world to supplement information in the text. Her research substantiates the classical work of Bartlett (1932), who found numerous distortions in the story recalls of English subjects who had read a tale of the Indians of the Northwest coast. The distortions made the story comply with the past experiences of the readers, suggesting that when misunderstandings or lapses in memory occur, readers reconstruct meaning based on their previous knowledge and experience.

The theoretical explanation for this view of reading as a constructive process is that as experiences and attitudes are assimilated they form cognitive structures. These cognitive structures, called *schemata* (Anderson, 1977; Rumelhart, 1980; Spiro, 1977), serve as a framework for storing information and for interpreting information implicit in the text. Thus, in Bartlett's study, the subjects' memory for a story from an unfamiliar culture was distorted because recall was based on and conformed with subjects' prior knowledge. When readers cannot exactly recall aspects of a story, they rely on previously formed schemata to reconstruct what might have occurred. According to Kintsch and van Dijk (1978), familiarity with the facts allows readers to make inferences and to fill in missing ideas. Readers recall not only what is stated, but what seems to follow.

Bransford and Franks (1971) demonstrated that college undergraduates acquired complete ideas from exposure to partial ideas and that subjects genuinely believed they had originally been presented with the entire idea, when in fact they had not. Research by Brown and colleagues (1977) confirms these findings. Subjects in their study later had difficulty distinguishing between their own story embellishments and the actual prose content. Therefore, substantial empirical data support the presence of schemata that provide a basis for comprehending, interpreting, and remembering discourse.

As the literature review has demonstrated, text comprehensibility cannot be considered a property of text alone, but one of text-reader interaction. Accordingly, to estimate text comprehensibility, you must have some estimate of the influence of textual factors, such as conventional text difficulty level and adjunct comprehension aids, and an estimate of cognitive factors, such as the reader's prior knowledge of the text topic. But word recognition skill, a second reader factor, also has a profound effect on reading comprehension performance.

Word Recognition Skill

A central claim by Perfetti (1977), Lesgold and Perfetti (1978), Perfetti and Lesgold (1977, 1979), Stanovich (1980), and Perfetti and Roth (1981) is that difficulty in word identification not only subtly retards reading comprehension, but also severely disrupts comprehension of text by interrupting the reader's ongoing train of thought. As explained by Samuels (1983), when a person is automatic at word recognition, little attention to decoding is required and the available attention can be used for comprehension. If too much attention is required to decipher words, there will not be enough processing capacity left for comprehension. Thus, accuracy and automaticity of word recognition facilitate reading comprehension. Having to expend effort at decoding in a word by word manner leaves the reader with little capacity for higher order reading and comprehension.

The punitive consequences that slow decoders' experience are explained in terms of the limited capacity of the short term memory that can retain only four to seven items at one time (Kintsch & Vipond, 1979). Because poor decoders require considerable processing space to unlock single words, the theory is that the same processing space cannot be used to store previously coded words or phrases. Consequently, antecedent words are lost from memory. In addition, the reader's capacity to call up existing schemata to predict upcoming information will likely be reduced. As a result, comprehension suffers. Once decoding skills function appropriately, the assumption is that processing capacity will be free and comprehension performance will improve. Readers then will be able to focus attention on the task of text comprehension, processing at deep structure levels as opposed to surface structure levels.

While observers would argue that some poor comprehenders have adequate word recognition skill, studies by Calfee and Drum (1978), Golinkoff and Rosinski (1976), and Perfetti and Hogaboam (1975) report that the apparent sufficiency in word recognition ability is nullified when speed is introduced as a dependent variable. Using third and fifth grade pupils as subjects, Perfetti and Hogaboam demonstrated that good comprehenders decode single words faster than do poor comprehenders. For common, high frequency words used in their study, differences in decoding times were slight but still significant. For nonwords conforming to English spelling patterns, however, decoding time differences were more pronounced. This suggests that less skilled readers expend too much effort on word identification and to enhance comprehension performance readers need to be more than accurate at word recognition. LaBerge and Samuels (1974) have labeled this additional dimension associated with word recognition as automaticity.

Both accuracy and automaticity of word recognition are important to reading comprehension. As a result, word recognition skill is a second inside the head cognitive factor likely to have a significant effect on text comprehensibility.

A New Approach to Predicting Readability

A study by Zakaluk (1985) found support for the theoretical constructs proposed in the foregoing review of the literature. She studied fifth grade students from twelve classrooms in urban, suburban, and rural schools. The students read passages ranging from 350 to 435 words taken from social studies and science-health texts. The difficulty of the passages ranged from grades four through seven (Fry, 1968). After reading, students answered open ended recall questions under various adjunct aid conditions. Results indicated that across four trials inside the head factors (word recognition automaticity, prior knowledge) combined with outside the head factors (passage difficulty as estimated by conventional readability measures and the use of adjunct comprehension aids) accounted for from 40 to 28 percent of the variance. Thus empirical data demonstrate that both inside and outside factors influence how well a given reader will comprehend a particular text.

Despite the importance of these factors, formulas currently used for estimating text readability fail to take a number of them into account. We therefore propose a new procedure to predict text difficulty using information from both inside and outside the head sources. We should be able to make better predictions about the ability of individual readers to comprehend particular texts by considering all elements: traditionally measured text factors (word difficulty and sentence length), adjunct aids (an outside the head factor), and the reader's prior knowledge and reading skills.

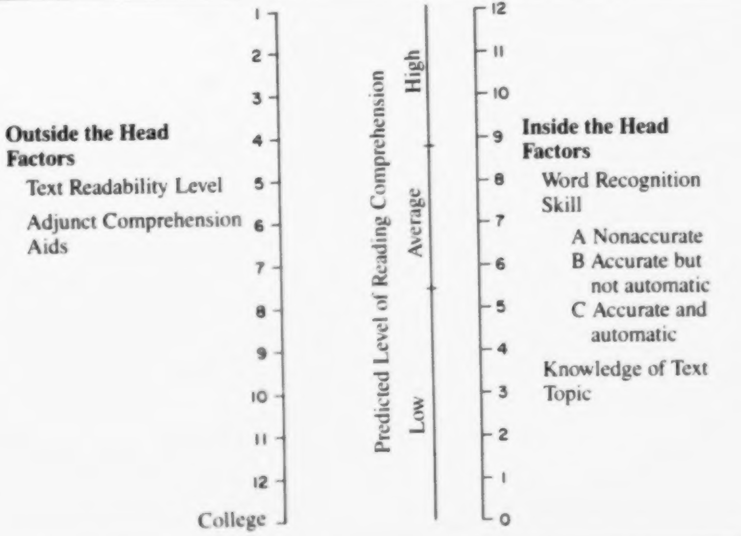
In order to simplify this process, we use a nomograph. A nomograph is a table that uses information from two sources to provide information about a third area of interest. A common application for a nomograph is to obtain an estimate of the percent of body fat. To do this, we measure skinfold thickness from two different parts of the body, such as the back of the upper arm and the rib cage.

We plot these two readings on two scaled vertical lines and then connect the points by using a ruler. Between the two outside lines is a third vertical line that serves as the predictor variable. Where the ruler crosses this predictor variable is the percent of body fat. We propose to estimate reading comprehension performance by the same process.

Using the Nomograph

Figure 1 shows the nomograph with three vertical lines. On the left a scale indicates outside the head factors that influence comprehensibility. These include text readability level and adjunct comprehension aids. Text readability ranges from grade one through college level. On the vertical line to the right, we find inside the head factors that influence comprehension. These include knowledge of text topic and word recognition skill. The center line indicates the extent to which an individual can comprehend the text in question.

Figure 1
A Nomograph for Predicting Text Comprehensibility


Outside the Head Factors
Text Readability Level
Adjunct Comprehension Aids
1
2
3
4
5
6
7
8
9
10
11
12
College
Predicted Level of Reading Comprehension
High
Average
Low
12
11
10
9
8
7
6
5
4
3
2
1
0
Inside the Head Factors
Word Recognition Skill
A Nonaccurate
B Accurate but not automatic
C Accurate and automatic
Knowledge of Text Topic

Inside the Head Factors

Measuring word recognition. There are three levels of word recognition skill. The lowest is the nonaccurate level, at which students experience difficulty in word recognition. The second is called accurate but not automatic. At this level, students devote most of their attention to decoding, leaving little capacity to consolidate overall meaning. At the third level, students are both accurate and automatic at word recognition. When students are automatic, word recognition requires minimal effort, thus allowing readers to focus their attention on obtaining meaning.

The simplest way to determine students' level of word recognition is to have them read orally from a 150 word passage that is at their grade placement level in terms of readability. Instruct students to read orally and to be able to tell what they remember when they finish. If students' word recognition accuracy is less than 95 percent, they are labeled nonaccurate. If students score above 95 percent in word recognition accuracy but experience difficulty retelling what they read, they fall into the accurate but not automatic category. If students achieve over 95 percent in terms of word recognition accuracy and can retell the gist of the passage satisfactorily,

Figure 2
Worksheet for Prior Knowledge Responses

Fossil Fuels ______________________________
Fossil Fuels ______________________________
Fossil Fuels ______________________________
Fossil Fuels ______________________________
Fossil Fuels ______________________________
Fossil Fuels ______________________________
Fossil Fuels ______________________________
Fossil Fuels ______________________________

The worksheet should contain about 25 lines with the stimulus words written at the beginning of each line.

they are both accurate and automatic. Nearly accurate oral reading and the ability to capture the gist of the selection indicate automaticity because we assume that simultaneous decoding and comprehension take place only when decoding is automatic.

Knowledge of text topic. We can use a word association technique to measure topic familiarity (Zakaluk, Samuels, & Taylor, 1986). A key word or phrase that embodies the main idea of the topic is chosen to serve as a stimulus. Students are required to write as many words or ideas as they can think of in association with the key word. Having the students use lined paper on which the stimulus word is printed at the beginning of each line (Figure 2) ensures that they continue to use the original word or phrase, not newly pro-

Figure 3
Instructions for Administering the Word Association Task

Introduction

This is a test to see how many words you can think of and write down in a short time.

You will be given a *key* word and you are to write as many *other* words as you can that the key word brings to mind.

The words you write may be things, places, ideas, events – whatever you think of when you see the key word.

Modeling and Chalkboard Demonstration

For example, think of the word *king*. (Write *king* on the chalkboard.) Some of the words or phrases that king brings to mind are queen/prince/palace/Charles/London/kingdom/England/ruler/kingfish/Sky King/of the road. Continue to brainstorm for other words. Add these to the chalkboard list. You may use two words, phrases, long words, or short words. Any idea is acceptable, no matter how many words.

Practice with Discussion

Work on practice sheets. *Kitchen* and *transportation* are two highly familiar topics. Following completion of the activity, clarify the task by sharing ideas and discussing any questions.

Reminders

The following reminders are given during practice and during the actual task.

1. No one is expected to fill in all the spaces on the page, but write as many words as you can think of in association with the key word.
2. Be sure to remember the key word while writing because the test is to see how many other words the key word calls to mind.
3. A good way to do this is to repeat the key word over and over to yourself as you write.

duced words, to generate ideas. Figure 3 gives sample directions for this task. Give the students three minutes to generate words and ideas.

Responses are scored with one point being awarded for each reasonable idea unit up to a maximum of ten points. No credit is granted for unreasonable associations, for example, the word *sandwich* in conjunction with the stimulus word *paper.* When generated words or phrases consist of a list that can be subsumed under a superordinate category, one point is given for the superordinate category and one point for all of the subordinate ideas. For example, if the key word is *farming* and a student lists the names of a series of crops such as wheat, barley, corn, rye, and oats, one point is given for the superordinate word *crops* and one point for the itemized products. In this case it is assumed that students have begun to use the generated words rather than the stimulus word as cues for producing

Outside the Head Factors

Readability level. To establish the difficulty level of the text, a readability formula such as Fry's (1968) may be used. If the text is one used in the classroom, such as a basal reader or a social studies text, the readability level is often the same as the grade level for which the text is designed. Be cautious in applying this rule, however, because different selections within the text may vary.

Adjunct comprehension aids. In addition to text readability level, another outside the head factor that influences comprehension is adjunct aids, such as statements of objectives or study questions located within the text or at the beginning or the end. These adjunct comprehension aids highlight important information and increase the depth of text processing. Their presence in the material adds to text comprehensibility.

We have described how to determine inside and outside the head factors. Now we indicate how to use these with the nomograph shown in Figure 1.

Plotting the Nomograph

To establish the outside the head factors, enter the scale at the text readability level. To determine a composite outside the head factor score, subtract a half point if a statement of objectives is present and another half point if study questions are present.

To determine the inside the head factors, for word recognition skill add zero if the student is nonaccurate (below 95 percent when substitutions, mispronunciations, word omissions, additions, and repetitions that involve two or more words are tallied from the oral reading); one point if the student is accurate but not automatic; and two points if the student is both accurate and automatic, as outlined in the following scale.

Points	*Word Recognition Level*
0	Nonaccurate
1	Accurate but not automatic
2	Accurate and automatic

For knowledge of text topic, add the score obtained from the word association task (maximum of ten points) to the word recognition score, thus establishing an overall score for plotting on the inside the head factor scale. Connect the plotted scores on the two outside scales with a ruler and read the predicted level of comprehension performance on the center scale. This will be either high, average, or low.

Figure 4 illustrates the application of the nomograph for text written at college levels. For outside the head factors, the text readability level is college and there are no adjunct aids. For inside the head factors, the student is automatic in terms of word recognition (two points) and generated five ideas on the word association task (five points). The student's inside the head score is thus 7. When the inside and outside the head figures are connected, the predicted level of comprehension as indicated on the center line is just below average.

Figure 4
Application of the Nomograph for College Level Text

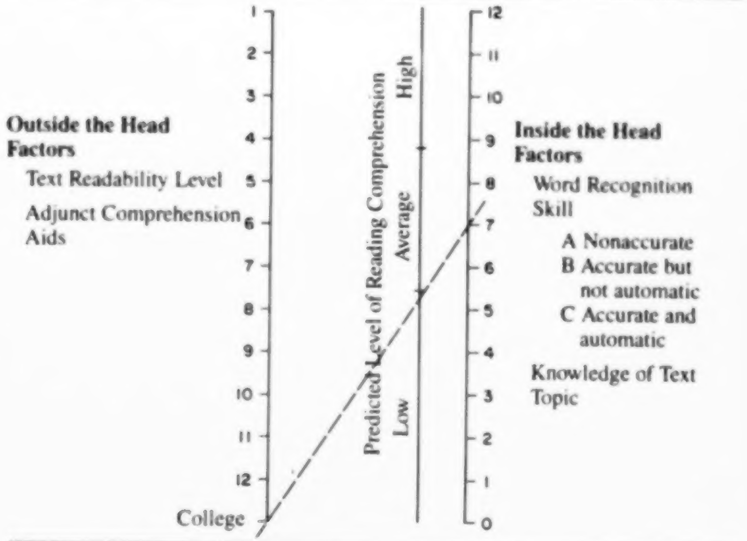

Outside the Head Factors
Text Readability Level
Adjunct Comprehension Aids
1
2
3
4
5
6
7
8
9
10
11
12
College
Predicted Level of Reading Comprehension
High
Average
Low
12
11
10
9
8
7
6
5
4
3
2
1
0
Inside the Head Factors
Word Recognition Skill
A Nonaccurate
B Accurate but not automatic
C Accurate and automatic
Knowledge of Text Topic

Figure 5
Application of the Nomograph for Primary Level Text

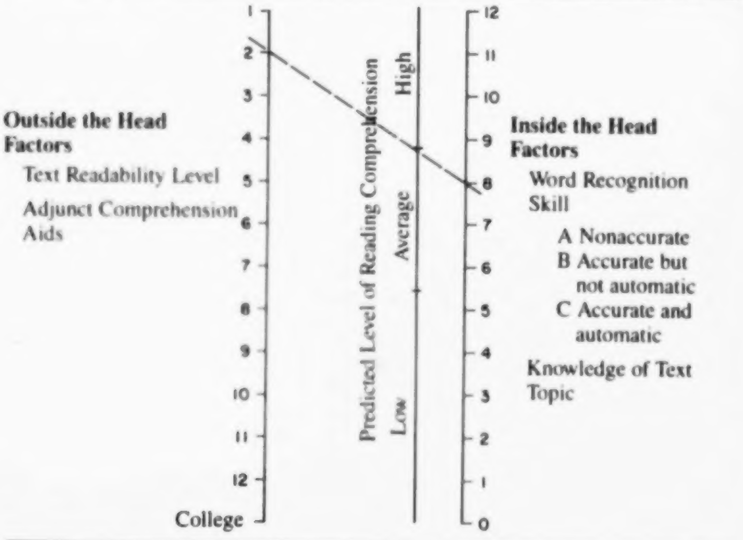
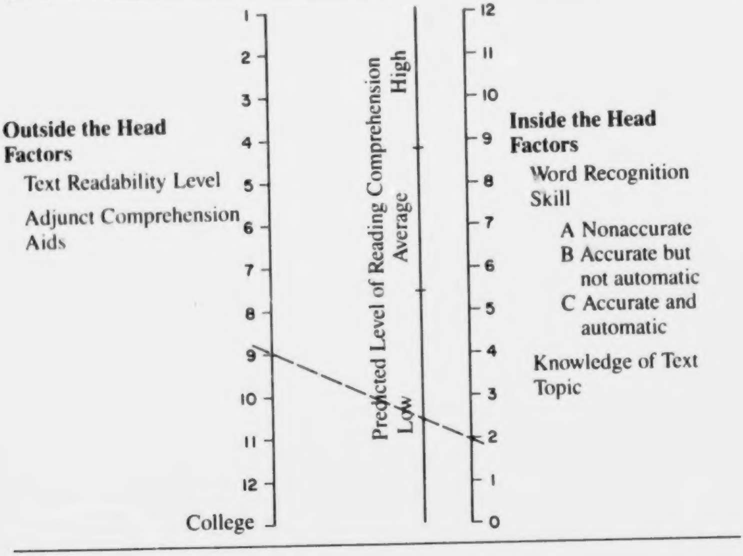

Outside the Head Factors
Text Readability Level
Adjunct Comprehension Aids
1
2
3
4
5
6
7
8
9
10
11
12
College
Predicted Level of Reading Comprehension
High
Average
Low
12
11
10
9
8
7
6
5
4
3
2
1
0
Inside the Head Factors
Word Recognition Skill
A Nonaccurate
B Accurate but not automatic
C Accurate and automatic
Knowledge of Text Topic

Figure 5 illustrates the application of the nomograph at primary levels. The outside the head factors show that the text readability level is grade two and there are no adjunct comprehension aids. The inside the head factors indicate that the student is accurate but not automatic at word recognition (one point) and obtained a score of 7 on the word association task (seven points). The inside the head factor score is therefore 8. The predicted level of comprehension for that student consequently is high average.

Our final example is of a student who is reading a tenth grade text that contains a statement of objectives as well as questions interspersed throughout the text. We enter the outside the head scale at grade ten and subtract one-half point for each of the adjunct comprehension aids (subtract because adjunct aids decrease the difficulty level of the text). The result gives us an outside the head composite score of 9. The student is automatic at word recognition (two points) but has no prior knowledge (zero points). The combined inside the head factor is therefore two, and the predicted comprehension performance is low, as indicated in Figure 6.

Figure 6
Application of the Nomograph for a Secondary Level Text


Outside the Head Factors
Text Readability Level
Adjunct Comprehension Aids
1
2
3
4
5
6
7
8
9
10
11
12
College
Predicted Level of Reading Comprehension
High
Average
Low
12
11
10
9
8
7
6
5
4
3
2
1
0
Inside the Head Factors
Word Recognition Skill
A Nonaccurate
B Accurate but not automatic
C Accurate and automatic
Knowledge of Text Topic

The nomograph was validated by comparing the predicted comprehension performance of hypothetical readers as estimated by this simple to use nomograph with a more detailed nomograph that was developed on the actual performance of the 253 fifth grade readers who were subjects in Zakaluk's study (1985) cited earlier. When the actual predictions using the complex nomograph were compared with the nomograph predictions derived from the nomograph presented in this chapter, there was a match on 30 of 36 comparisons (r = .93). In other words, when hypothetical cases were generated to test the degree of overlap between comprehension predictions made with the two nomographs, the validity of the simple to use nomograph was upheld.

Conclusions

Students make optimal learning gains when instructional text matches their reading achievement level. Current measures to estimate text difficulty are inadequate because they consider only one source of information—that contained on the printed page. In addition to the outside the head factors of text readability levels and the use of adjunct comprehension aids, inside the head factors also must be considered in predicting the difficulty level of a particular text for a particular reader. These inside the head factors include knowledge of text topic and degree of reading fluency.

This nomograph has taken the prediction of text readability one step further, bringing in powerful variables that influence text comprehension. Thus, the teacher is able to make better estimates about which students will comprehend with ease and which will require extra attention and be able to adjust instructional approaches accordingly, spending more time on developing word familiarity or building and activating prior knowledge. The nomograph draws readability out of the stage of behaviorism where only outside the head factors were examined, bringing the concept into the realm of cognitive psychology where both inside and outside the head factors are examined to predict reading performance.

References

Adams, M.J., and Collins A. A schema theoretic view of reading. In R.O. Freedle (Ed.), *New directions in discourse processing*, volume two. Norwood, NJ: Ablex, 1979.

Anderson, R.C. The notion of schemata and the educational enterprise: General discussion of the conference. In R.C. Anderson, R.J. Spiro, and W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum, 1977.

Anderson, R.C., and Biddle, W.B. On asking people questions about what they are reading. In G. Bower (Ed.), *Psychology of learning and motivation*, volume 9. New York: Academic Press, 1975.

Anderson, T.H. Study strategies and adjunct aids. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum, 1980.

Andre, T. Does answering higher level questions while reading facilitate productive learning? *Review of Educational Research*, 1979, *49*, 280-318.

Armbruster, B.P., and Anderson, T.H. *Content area textbooks*. Technical Report No. 23. Champaign, IL: Center for the Study of Reading, University of Illinois, 1981.

Ausubel, D.P. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 1960, *51*, 267-272.

Balow, I.A. Does homogeneous grouping give homogeneous groups? *Elementary School Journal*, 1962, *53*, 28.

Bartlett, F.C. *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press, 1932.

Betts, E.A. *Foundations of reading instruction*. New York: American Book, 1957.

Blachowicz, C.L. Semantic constructivity in children's comprehension. *Reading Research Quarterly*, 1977, *13*, 188-199.

Bond, G.L., and Wagner, E.B. *Teaching the child to read*, fourth edition. New York: Macmillan, 1966.

Bormuth, J.R. Measurement of readability. In L.C. Deighton (Ed.), *The encyclopedia of education*, volume 7. New York: Crowell-Collier, 1971.

Bransford, J.D., and Franks, J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, *2*, 330-350.

Bransford, J.D., and Johnson, M. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, *1*, 717-726.

Bransford, J.D., and McCarrell, N.S. A sketch of a cognitive approach to comprehension. In W.B. Wiemer and D.S. Palermo (Eds.), *Cognition and the symbolic processes*. Hillsdale, NJ: Erlbaum, 1974.

Brown, A.L., Smiley, S.S., Day, J.D., Townsend, M.A., and Lawton, S.C. Intrusion of a thematic idea in children's comprehension and retention of stories. *Child Development*, 1977, *48*, 1454-1466.

Calfee, R.C., and Drum, P.A. Learning to read: Theory, research, and practice. *Curriculum Inquiry*, 1978, *8*, 183-249.

Chiesi, H.L., Spilich, G.J., and Voss, J.F. Acquisition of domain related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 257-273.

Coleman, E.B. Developing a technology of written instruction: Some determiners of the complexity of prose. In E.L. Rothkopf and R.E. Johnson (Eds.), *Verbal learning research and the technology of written instruction*. New York: Teachers College, Columbia University, 1971.

Dale, E., and Chall, J. A formula for predicting readability. *Educational Research Bulletin*, 1948, *27*, 11-20.

Dale, E., and Tyler, R.W. A study of the factors influencing the difficulty of reading materials for adults of limited reading ability. *Literary Quarterly*, 1934, *4*, 384-412.

Dooling, D., and Lachman, R. Effect of comprehension on retention of prose. *Journal of Experimental Psychology*, 1971, *88*, 216-222.

Duchastel, P.C., and Merrill, P.F. The effects of behavioral objectives on learning: A review of empirical studies. *Review of Educational Research*, 1973, *43*, 53-69.

Duell, O.K. Effects of type of objective level of test questions, and the judged importance of tested materials upon posttest performance. *Journal of Educational Psychology*, 1974, *66*, 225-232.

Dunkeld, C. *The validity of the informal reading inventory for the designation of instructional reading levels*. Unpublished doctoral dissertation, University of Illinois, 1970.

Faw, H.W., and Waller, T.G. Mathemagenic behaviors and efficiency in learning from prose materials: Review, critique, and recommendations. *Review of Educational Research*, 1976, *46*, 691-720.

Finn, P.J. The use of the relative frequency coefficient to identify high information words and to measure readability. In G.H. McMinch and W.D. Miller (Eds.), *Reading: Convention and inquiry*. Clemson, SC: National Reading Conference, 1975.

Flesch, R. A new readability yardstick. *Journal of Applied Psychology*, 1948, *32*, 221-233.

Frase, L.T. Learning from prose material: Length of passage, knowledge of results, and position of questions. *Journal of Educational Psychology*, 1967, *58*, 266-272.

Fry, E. A readability formula that saves time. *Journal of Reading*, 1968, *11*, 513-516; 575-578.

Golinkoff, R.N., and Rosinski, R.R. Decoding, semantic processing, and reading comprehension skill. *Child Development*, 1976, *47*, 252-258.

Gray, W.S., and Leary, B.E. *What makes a book readable?* Chicago: University of Chicago Press, 1935.

Jenkins, J.R., and Deno, S.L. Influence of knowledge and type of objectives on subject matter learning. *Journal of Educational Psychology*, 1971, *62*, 67-70.

Johnson, M.S., Kress, R.A., and Pikulski, J.J. *Informal reading inventories*, second edition. Newark, DE: International Reading Association, 1986.

Kaplan, R. Effects of learning prose with part versus whole presentations of instructional objectives. *Journal of Educational Psychology*, 1974, *66*, 787-792.

Kintsch, W., Kozminsky, E., Streby, W.J., McKoon, G., and Keenan, J.M. Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 196-214.

Kintsch, W., and Miller, J.R. Readability: A view from cognitive psychology. In J. Flood (Ed.), *Understanding reading comprehension*. Newark, DE: International Reading Association, 1984.

Kintsch, W., and van Dijk, T.A. Toward a model of text comprehension and prediction. *Psychological Review*, 1978, *85*, 363-394.

Kintsch, W., and Vipond, D. Reading comprehension and readability in educational practice and psychological theory. In L. Nilsson (Ed.), *Proceedings of the conference on memory*. Hillsdale, NJ: Erlbaum, 1979.

LaBerge, D., and Samuels, S.J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293-323.

Lesgold, A.M., and Perfetti, C.A. Interactive processes in reading comprehension. *Discourse Processes*, 1978, *1*, 323-336.

Lively, B.A., and Pressey, S.L. A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision*, 1923, *9*, 389-398.

Lorge, I. Predicting reading difficulty of selections for children. *Elementary English Review*, 1939, *16*, 229-233.

Lorge, I. Predicting readability. *Teachers College Record*, 1944, *45*, 404-419.

Manelis, L., and Yekovich, F.R. Repetitions of propositional arguments in sentences. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 301-312.

Marshall, N. Research: Reading and comprehensibility. *Journal of Reading*, 1979, *32*, 542-544.

McCall, W.A., and Crabbs, L.M. *Standard test lessons in reading*, Books 2, 3, 4, 5. New York: Bureau of Publications, Teachers College, Columbia University, 1926.

McConnell, G.R. Readability formulas as applied to college economics text books. *Journal of Reading*, 1982, *26*, 14-17.

Nelson-Herber, J. Readability: Some cautions for the content area teacher. In W.J. Harker (Ed.), *Classroom strategies for secondary reading*, second edition. Newark, DE: International Reading Association, 1985.

Pearson, P.D. The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. *Reading Research Quarterly*, 1974, *10*, 155-192.

Pearson, P.D., Hansen, J.D., and Gordon, C. The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior*, 1979, *11*, 201-209.

Perfetti, C.A. Language comprehension and fast decoding: Some psycholinguistic prerequisites for skilled reading comprehension. In J.T. Guthrie (Ed.), *Cognition, curriculum, and comprehension*. Newark, DE: International Reading Association, 1977.

Perfetti, C.A., and Hogaboam, T.W. The relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology*, 1975, *67*, 461-469.

Perfetti, C.A., and Lesgold, A.M. Discourse comprehension and sources of individual differences. In M. Just and P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Erlbaum, 1977.

Perfetti, C.A., and Lesgold, A.M. Coding and comprehension in skilled reading and implications for reading instruction. In L.B. Resnick and P. Weaver (Eds.), *Theory and practice of early reading*. Hillsdale, NJ: Erlbaum, 1979.

Perfetti, C.A., and Roth, S. Some of the interactive processes in reading and their role in reading skill. In A. Lesgold and C. Perfetti (Eds.), *Interactive processes in reading*. Hillsdale, NJ: Erlbaum, 1981.

Rothkopf, E.Z. Learning from written instructive materials: An exploration of the control of inspection behavior by test like events. *American Educational Research Journal*, 1966, *3*, 241-249.

Rothkopf, E.Z., and Bisbicos, E. Selective facilitative effect of interspersed questions on learning from written material. *Journal of Educational Psychology*, 1967, *58*, 56-61.

Rothkopf, E.Z., and Kaplan, R. Exploration of the effect of density and specificity of instructional objectives on learning from text. *Journal of Educational Psychology*, 1972, *63*, 295-302.

Rumelhart, D.E. Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce and W.F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum, 1980.

Rumelhart, D.E., and Ortony, A. The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, and W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum, 1977.

Samuels, S.J. Introduction to theoretical models of reading. In W. Otto, C. Peters, and N. Peters (Eds.), *Reading problems: A multidisciplinary perspective*. Reading, MA: Addison-Wesley, 1977.

Samuels, S.J. Diagnosing reading problems. *Topics in Learning and Learning Disabilities*, 1983, *2*, 1-11.

Scarborough, R.F., Bruns, R.F., and Frazier, A. Improvement of reading through ability level assignments. *School Review*, 1957, *65*, 474-480.

Spiro, R.J. Remembering information from text: The state of schema approach. In R.C. Anderson, R.J. Spiro, and W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum, 1977.

Stanovich, K. Toward an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 1980, *16*, 32-71.

Stevens, K.C. Readability formulae and McCall-Crabbs standard test lessons in reading. *The Reading Teacher*, 1980a, *33*, 413-415.

Stevens, K.C. The effect of background knowledge on the reading comprehension of ninth graders. *Journal of Reading Behavior*, 1980b, *12*, 151-154.

Taylor, B.M. Good and poor readers' recall of familiar and unfamiliar text. *Journal of Reading Behavior*, 1979, *11*, 375-380.

Vogel, M., and Washburne, C. An objective method of determining grade placement of children's reading material. *Elementary School Journal*, 1928, *28*, 373-381.

Zakaluk, B.L. *Toward a new approach to predicting text comprehensibility using inside- and outside-the-head information and a nomograph*. Unpublished doctoral dissertation, University of Minnesota, 1985.

Zakaluk, B.L., Samuels, S.J., and Taylor, B. A simple technique for estimating prior knowledge: Word association. *Journal of Reading*, 1986, *30*, 56-60.